İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

TÜRKÇE İÇİN BİÇİMBİRİMSEL BELİRSİZLİK GİDERİCİ

YÜKSEK LİSANS TEZİ

Müh. Z. İlknur KARADENİZ

Anabilim Dalı : BİLGİSAYAR MÜHENDİSLİĞİ

Programı : BİLGİSAYAR MÜHENDİSLİĞİ

OCAK 2007

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

TÜRKÇE İÇİN BİÇİMBİRİMSEL BELİRSİZLİK GİDERİCİ

YÜKSEK LİSANS TEZİ
Müh. Z.İlknur KARADENİZ
(504041547)

Tezin Enstitüye Verildiği Tarih : 27 Aralık 2007
Tezin Savunulduğu Tarih : 30 Ocak 2007

Tez Danışmanı :	**Prof.Dr. Eşref ADALI**
Diğer Jüri Üyeleri	**Prof.Dr. Oya KALIPSIZ (Yıldız Teknik Ü.)**
	Yrd.Doç.Dr. Zehra ÇATALTEPE (İ.T.Ü.)

OCAK 2007

ÖNSÖZ

Öncelikle, danışmanım Prof.Dr.Eşref Adalı'ya tezin başlangıç aşamasından sonuna kadar verdiği destek, motivasyon ve bana gösterdiği sonsuz sabrı için teşekkür ederim. İ.T.Ü.Doğal Dil İşleme Grubu üyesi Araş.Gör. Gülşen Cebiroğlu Eryiğit'e gösterdiği sıcak ilgi, tezimle ilgili değerli görüşleri ve umutsuzluğa kapıldığım her anda verdiği cesaret için minnettarım. Oda arkadaşım Araş.Gör. Berk Canberk başta olmak üzere diğer tüm araştırma görevlisi arkadaşlarıma ve bütün İ.T.Ü. Bilgisayar Bölümü mensuplarına, eğitim ve özel yaşamımda her konuda beni destekleyen aileme ve sevgisini benden esirgemeyen Dr. Emre Erol'a teşekkürü bir borç biliyorum.

Aralık, 2006 **Z. İlknur KARADENİZ**

İÇİNDEKİLER

KISALTMALAR **V**
TABLO LİSTESİ **Vİ**
ŞEKİL LİSTESİ **Vİİ**
ÖZET **Vİİİ**
SUMMARY **İX**
1. GİRİŞ **1**
1.1 Benzer Çalışmalar 1
1.1.1 Biçimbiririmsel Etiketleyiciler 2
1.1.2 Biçimbirimsel Belirsizlik Gidericiler 4
1.2 Tezin Ana Hatları 7
1.3 İçerik 8

2. TEMEL KAVRAMLAR **10**
2.1 Doğal Dil İşleme 10
2.2 Biçimbirimsel Belirsizlik 12
2.3 Türkçe'nin Biçimbirimsel Yapısı 12
2.4 Derlem 14

3. TÜRKÇE İÇİN KURAL TABANLI BİÇİMBİRİMSEL BELİRSİZLİK GİDERİCİ **15**
3.1 Türkçe'deki Biçimbirimsel Belirsizlikler 15
3.1.1 Gövde Belirsizlikleri 15
3.1.2 Ana Etiket Belirsizlikleri 16
3.1.3 Çekim Grubu Belirsizlikleri 17
3.1.4 Diğer Belirsizlik Analizleri 18
3.1.5 Kalıplar 18
3.2 Türkçe'de En Sık Kullanılan Belirsiz Sözcükler 19
4. GEREKSİNİMLER VE PROGRAMLAMA ORTAMI **29**
5. SINAMA **30**
6. SONUÇLAR VE TARTIŞMA **33**

KAYNAKLAR **34**

EK A **36**

EK B **51**

ÖZGEÇMİŞ **67**

KISALTMALAR

Acc	**:**-i hali
AÖA	**:**Açgözlü Önden Eklemeli Algoritma
A1sg	**:**1.Tekil Şahıs Kişi Eki
A2sg	**:**2.Tekil Şahıs Kişi Eki
A3sg	**:**3.Tekil Şahıs Kişi Eki
A1pl	**:**1.Çoğul Şahıs Kişi Eki
A2pl	**:**2.Çoğul Şahıs Kişi Eki
A3pl	**:**3.Çoğul Şahıs Kişi Eki
Dat	**:**-e hali
DDİ	**:**Doğal Dil İşleme
Loc	**:**-de hali
Nom	**:** Yalın hali
P1sg	**:**1.Tekil Şahıs İyelik Eki
P2sg	**:**2.Tekil Şahıs İyelik Eki
P3sg	**:**3.Tekil Şahıs İyelik Eki
P1pl	**:**1.Çoğul Şahıs İyelik Eki
P2pl	**:**2.Çoğul Şahıs İyelik Eki
P3pl	**:**3.Çoğul Şahıs İyelik Eki

TABLO LİSTESİ

Sayfa No

Tablo 1.1: Macarca için Brill'in Etiketleyicisinin Başarımı 4
Tablo 1.2: Farklı Modellerin Başarımları 5
Tablo 1.3: Örnek Karar Listesi.....6
Tablo 2.1: Örnek Fiil Çekimi 13
Tablo 2.2: -ecek/-acak Ekiyle Oluşan Sözcükler 13

ŞEKİL LİSTESİ

Sayfa No

Şekil 1.1: Biçimbirimsel Belirsizlik Dağılımları 8
Şekil 3.1: Gövde Belirsilikleri 16
Şekil 3.2: Ana Etiket Belirsizlikleri 17
Şekil 3.3: Çekim Grubu Belirsizlikleri: 18
Şekil 3.4: En Sık Kullanılan Belirsizlik Sözcükler 19

TÜRKÇE İÇİN BİÇİMBİRİMSEL BELİRSİZLİK GİDERİCİ

ÖZET

Doğal Dil İşlemenin kapsamında olan, özet çıkartma, metin sınıflandırması, diller arası çeviri ve anlamsal çalışmalarda bir sözcüğün niteliğinin belirlenmesi önemli bir konudur. Bir sözcüğün niteliği demek, en temel anlatımla o sözcüğün isim mi fiil mi sıfat mı olduğudur. Biraz ayrıntıya inildiğinde, daha alt özelliklerin de belirtilmesi gerekir. Sözgelimi bir fiilin zamanı ve şahsının belirtilmesidir.

Türkçe bitişken bir dil olması nedeniyle, sözcükler kök, gövde ve eklerden oluşur. Türkçe ön ek açısından çok kısıtlı olmasına karşın, son ekler konusunda olağanüstü zengindir. Bu eklerin 300 civarında bulunduğu ve bunların yarısının yoğun bir şekilde kullanıldığı bilinmektedir. Türkçe'nin yeni sözcük türetmesine büyük katkı sağlayan bu özelliği, biçimbirimsel çözümlemede sorunlara neden olmaktadır. Bir Hint-Avrupa ailesi dilinde bir sözcüğün biçimbirimsel analizi sonunda belirsizlik oranı az iken, Türkçe'de bir sözcük için 1,8 çözüm bulunmaktadır.

Bu tezin amacı Türkçe için biçimbirimsel çözümleyicinin ürettiği sonuçlardan doğru olanının saptanmasıdır. Bu amaçla önce Türkçe'deki sözcüklerin belirsizlik dağılımları çıkartıldı. Ardından sözcükler belirsizlik niteliklerine göre kümelendi. Her belirsizlik türü için kurallar yazarak belirsizlikler giderilmiştir. Gerçeklediğimiz çalışmanın başarımı sınanmış ve başarımının %82,6 olduğu görülmüştür.

MORFOLOGICAL DISAMBIGUATOR for TURKISH

SUMMARY

It is an important issue in summarization, text - categorization and semantic related works ,which are the study areas of Natural Language Processing , to detect the properties of a word .The property of a word basically means whether the word is an adjective, a noun or a verb, etc. More precisely, it is not that simple. The sub-categories (for instance, the tense and personal properties of a verb) should also be analayzed.

Because of the agglutinative nature of Turkish, Turkish words are made of roots, stems and affixes. Turkish has restricted prefixes but there are roughly 300 suffixes. It is known that the half of these suffixes are widely used. Although this feature helps new word formations, also produces new morphological problem with it. While, in a n Euraopean language, the morphological analaysis of a word produces a few results for a word, this number is 1,8 for a Turkish word.

The aim of this thesis is selecting the true parse of a word in a given context from the results produced by a morphological analyzer. For achieving this goal, firstly the morpholgical ambiguities in Turkish is classified and the distribution of the ambiguiteis are analyzed..Then, for all categories, spesific diambiguation rules are written to get rid of the morphologiacl ambiguities. The performance of our work is tested and the sucess is measured as 82.6%.

1.. GİRİŞ

Doğal Dil İşleme (DDİ) yapay zekanın bir alt dalıdır. Amaç bilgisayarların ana dili anlamasını, insanların bilgisayarla olan ilişkilerinde ana dillerini kullanabilmesini sağlamaktır.

Biçimbirim doğal dil işlemenin sözcük yapısıyla ilgili düzeyidir. Bu düzeyde biçimbirimsel çözümleyici kullanılarak bir sözcüğün biçimbirimsel özellikleri bulunabilir.

Sözcükleri biçimbirimsel çözümleyiciden geçirdiğimizde, her zaman bir tek sonuç ortaya çıkmayabilir.Birden fazla çözümleme sonucu ortaya çıktığında amaç sözcüğün geçtiği cümle içerisindeki anlamının bulunmasıdır.Bu da biçimbirimsel belirsizlik gidericilerin görevidir.

Biçimbirimsel belirsizlik giderme doğal dil işleme uygulamalarında büyük önem taşımaktadır. İstatistiksel, kural tabanlı ve dönüşüm tabanlı olmak üzere çok farklı teknikler kullanılarak çeşitli dillerdeki biçimbirimsel belirsizlik giderilmeye çalışılmıştır.

Türkçe için de yapılmış birçok uygulama bulunmaktadır.Bunlarla ilgili bilgiler 1.1.2 de ayrıntılı olarak verilmiştir.

Bu tez çalışmasının amaç ve kapsamı şöyledir.

Bir biçimbirimsel çözümleyicinin çözümlemesi sonunda ortaya çıkan sonuçlardan hangisinin doğru olduğunu saptanacaktır

En doğru sonucu bulmak için sadece ele alınan sözcük değil, komşu sözcüklerle olan ilişkileri de ele alınacaktır

1.1 Benzer Çalışmalar

Bu bölümde, tezle konumuza yakın olan çalışmalar değerlendirilecektir. Çalışmamızla ilgili olarak konuları iki küme altında toplayabiliriz:

- Biçimbirimsel Etiketleyiciler
- Biçimbirimsel Belirsizlik Gidericiler

1.1.1 Biçimbiririmsel Etiketleyiciler

Bir cümle içerisindeki her sözcüğün en uygun sıfat, isim, fiil veya benzeri ile işaretlenmesine biçimbirimsel etiketleme denir. Bir başka deyişle cümle içnde yer alan sözcüklerin nitelikleriyle etiketlenmesidir.

Bu bölümde daha önce yapılmış olan biçimbirimsel etiketleyiciler tanıtılacaktır.

1.1.1.1 Brill'in Etiketleyicisi

[1] ve [2] deki çalışmalarında; Brill, eğitim derleminden kullanıcı etkileşimi olmadan kuralları otomatik olarak çıkaran kural tabanlı bir etiketleyici önermiştir. Çok küçük bir dilbilgisi sisteme entegre edilmiştir, ancak bu bilgi dile bağımlı değildir. Gerekli olan tek şey eğitim derlemi olarak kullanılmak üzere önceden elle doğruları işaretlenmiş küçük ölçekli bir derlemdir. Bu eğitim derlemi etiketleyiciye girdi olarak verilir. Sistem böylelikle eğitim derleminden biçimbirimsel ve bağlamsal bilgileri türetip bir sözcüğü en uygun etiketiyle nasıl etiketleyeceğini öğrenir. Eğitim işlemi bittikten sonra, etiketleyici yeni, işaretlenmemiş bir derlemi etiketlemek için kullanılanabilir. Bu çalışmanın başarımı İngilizce için %97 olarak yazarı tarafından açıklanmıştır.

Brill'in etiketleyicisi melez bir etiketleyicidir. Çünkü ilk olarak istatistiksel tekniklerle eğitim derleminden gerekli bilgiyi elde edip, daha sonra istatistiksel yanlışlıklardan meydana gelen hataları azaltmak için otomatik kurallar üretir.

Etiketleyici ilk olarak, sözcüklerin geçtiği bağlama bakmaksızın derlemde geçme sıklığına göre sözcüklere en uygun etiketi atar. Bu başlangıç etiketleyicisinin başarımını artırmak için , bağlama bağlı olmayan iki aşama uygulanır. Bunlardan ilki, eğitim derleminde bulunmayan büyük harfli sözcüklerin özel isim olarak etiketlenmesidir. İkinci aşama ise, eğitim derleminde bulunmayan sözcüklere son üç harfleri aynı olan sözcüklere en sıklıkla atanan etiketin atanmasıdır. Örneğin, derlemimizde sonu "*ous"* ile biten sözcüklerimizin çoğu sıfat olarak etiketlendiyse "*blahblahous"* da aynı şekilde sıfat olarak etiketlenecektir (Örneğin, Türkçe'de de "lı" eki ile biten sözcükler çoğunluka sıfatsa, derlemimizde bulunmayan arabalı sözcüğü sıfat olarak etiketlenir).

Bu basit algoritma %92,1 oranında doğru etiketleme yapmıştır. Başarımı daha da artırmak için bir takım yamalar uygulanmıştır. Başlangıç etiketleyicisi derlemin %90ında eğitilmiştir. Geriye kalan derlemin %5i yama yapma aşaması için, diğer %5i ise sınama için ayrılmıştır.

Başlangıç etiketleyicisi, derlemimizin %90lık kısmında eğitildikten sonra %5lik yama derlemini etiketlemek için kullanılır. Yama derlemindeki sözcükler başlangıç etiketleyicisiyle etiketlendikten sonra, sözcüklerin doğru etiketleriyle karşılaştırılarak etiketleme hatalarından oluşan bir liste oluşturulur. Bu liste <$etiket_a$, $etiket_b$, sayı>

şeklinde bir üçlü içerir. Bu üçlü $etiket_b$ olması gerekirken $etiket_a$ olarak etiketlenen sözcüklerin sayısını ifade eder.

Yamalar kullanılarak bu yanlış etiketlendirmeler düzeltilmeye çalışılır. Yamalar şu şekillerde karşımıza çıkabilir:

$etiket_a$ yı $etiket_b$ ye değiştir eğer:

- Bir önceki(sonraki) sözcük z olarak etiketlendiyse
- İki önceki(sonraki) sözcük z olarak etiketlendiyse
- İki önceki(sonraki) sözcüklerden biri z olarak etiketlendiyse
- Üç önceki(sonraki) sözcüklerden biri z olarak etiketlendiyse
- Önceki sözcükz olarak ve sonraki sözcükde w olarak etiketlendiyse
- Bir önceki(sonraki) sözcük z olarak ve iki önceki(sonraki) sözcükde w olarak etiketlendiyse
- Sözcükbüyük harfle yazıldıysa(yazılmadıysa)
- Önceki sözcükbüyük harfle yazıldıysa(yazılmadıysa)

Her bir <$etiket_a$, $etiket_b$, sayı> üçlüsü ve uygulanan yama için, yamanın uygulanmasıyla hatalı etiketlerdeki azalma hesaplanır. Yalnız burada dikkat edilmesi gereken yamanın uygulanmasıyla yeni hataların da ortaya çıkabileceğidir. Dolayısıyla yamanın meydana çıkardığı yeni hatalar da hesaplamaya katılmalıdır.

Örneğin, başlangıç etiketleyicisi yama için kullanılan derlemi etiketlerken, 159 kelmeyi isim olması gerekirken fiil olarak etiketledi diyelim.“İki önceki sözcüklerden biri belirteç ise sözcüğün etiketini fiilden isime çevir” yaması uygulanırsa, bu 159 hatanın 98 ini düzeltir. Ancak bu sefer de gerçekten fiil olması gereken 18 sözcüğü de isime dönüştürür. Dolayısıyla bu yamanın uygulanmasıyla 80 hata düzeltilmiştir diyoruz.

Yamalardan en yüksek gelişmeyi sağlayan, yama listesine eklenir. Başarımı yüksek olan yama, yama derleminin etiketlenmesinde kullanılır.Bu aşama bu şekilde devam eder.

Yamaların bulunması işleminden sonra, yeni bir sınama derlemi etiketlenebilir. Öncelikle başlangıç etiketleyicisi kullanılarak etiketlenir. Daha sonra hata oranını azaltmak için sırasıyla yamalar uygulanır.

Bu yöntem İngilizce için çok iyi sonuç vermesine rağmen, Türkçe gibi bitişken diller için uygun değildir. Çünkü İngilizce’de verilen bir sözcükten türetilebilecek yeni sözcük sayısı sınırlıyken, Türkçe’de bir sözcüğün alabileceği ek sayısı sınırsız olduğundan, türetilebilecek sözcük sayısı da sınırsızdır. Sözcük sayısının sınırsız olmasından kaynaklı, etiket kümesi de sınırsızdır. Bu nedenle her hangibir biçimbirimsel çözümleyici

kullanmadan, otomatik çıkartılan kuralların uygulandığı bu yöntem Türkçe bir sözcüğün doğru sonucunun bulunması için iyi bir yöntem değildir.Yaptığımız deneyler sonuçaları bu değerlendirmemizin doğru olduğunu göstermiştir.

1.1.1.2 Macarca için Geliştirilmiş Brill'in Etiketleyicisi

[3] ve [4] deki çalışmalarda, bir önceki kısımda anlatılan Brill'in etiketleyicisi Macarca için aynen uygulanmıştır. Macarca da Türkçe gibi zengin bir biçimbirimsel yapıya sahip, bitişken bir dildir. Tablo in ilk satırında da görüldüğü gibi Macarca gibi zengin biçimbirime sahip bitişken diller için Brill'in etiketleyicisi iyi sonuçlar vermemiştir. Ancak kullanılan yama iskeletlerinde bir takım ufak değişiklikler yapılarak bu yöntemden alınan sonuçlar iyileştirilmiştir. Yapılan iyileştirmeler sonucundaki başarım artışı aynı tablonun ikinci satırında gözlemlenebilinir.

Tablo 1.1: Macarca için Brill'in Etiketleyicisinin Başarımı

	Sadece Ana Etiket	**Ana Etiket ve Doğru Alt Etiket**	**Ana Etiket ve Doğru Olmayan Alt Etiket**
Orjinal	%86,48	%85,98	%90,61
Geliştirilmiş	%95,53	%91,94	%97

1.1.2 Biçimbirimsel Belirsizlik Gidericiler

Bir sözcüğün biçimbirimsel çözümleme sonunda birden fazla çözümlemesi ortaya çıkabilir. Bu durumda bu sözcük için biçimbirimsel belirsizlik vardır denilir. Bu çözümlemelerden sözcüğün içinde bulunduğu bağlam için doğru olanının bulunmasına biçimbirimsel belirsizlik giderme adı verilir.

Bu bölümde daha önce yapılmış olan biçimbirimsel belirsizlik gidericiler tanıtılacaktır.

1.1.2.1 Bitişken Diller için İstatistiksel Biçimbirimsel Belirsizlik Giderme

[5] deki çalışmada, Türkçe ve Türkçe gibi bitişken özellikli dillerdeki yazıların biçimbirimsel belirsizliğinin giderilmesi için istatistiksel modeller önerilmiştir. Daha önce de bahsedildiği gibi Türkçe'nin bitişken özelliğinden dolayı olası etiket kümesi çok geniştir. Bu probleme çözüm olarak, bu çalışmada biçimbirimsel etiketlerin her biri çekim özelliklerini içerecek şekilde çekim gruplarına (IG) ayrılmıştır. Çalışmada, çekim grupları ve kökler göz önünde bulundurularak her bir etiketin olasılığı trigram istatistiksel modeller kullanılarak hesaplanır.

Biçimbirimsel belirsizlik giderme W = w_1, w_2 , …., w_n şeklinde verilen bir dizi sözcüğünT = t_1, t_2 , …., t_n şeklinde biçimbirimsel çözümlemelerle eşleşmesi problemidir. Bu çalışmada, Hidden Markov modeli kullanılarak, verilen sözcüklerin biçimbirimsel çözümlemelerinin dağılımı modellenmiş ve daha sonra da P(T|W)yi maksimize eden T değişkeni bulunmaya çalışılmıştır. P(T|W) maksimize edilmeye çalışılırken üç farklı model geliştirilmiş ve sınanmıştır. Bu modeller şunlardır:

MODEL1: Bir çekim grubunun sözcükde bulunma olasılığı, ondan önceki sözcüklerin son çekim gruplarına bağlıdır. Bu model, ilgili çekim grubunun varlığını o sözcükdeki diğer çekim gruplarından bağımsız olarak değerlendirmiştir.

MODEL2: Bir çekim grubunun sözcükde bulunma olasılığı, ondan önceki sözcüklerin son çekim gruplarına ve aynı sözcükdeki ilgili çekim grubundan önceki çekim grubuna bağlıdır.Yani bu modelde morfotaktik özellikler de göz önünde bulundurulmuştur.

MODEL3: Bu model Model 2 gibidir. Buradaki tek fark, bir çekim grubunun sözcükde bulunma olasılığının, ondan önceki sözcüklerin son çekim gruplarına bağlı olmamasıdır.

Bu üç modelden sözcüklerin kendi içlerindeki morfotaktik bağlantılarını göz önünde bulundurmayan model (Model 1) en iyi sonucu vermiştir. En iyi sonucu veren trigram tabloda gösterilmiştir.

Tablo 1.2: Farklı Modellerin Başarımları

Eğitim Verisi	**Model1**	**Model 1 Bigram**	**Model 2**	**Model3**
Belirli Sözcükler	88,21%	89,06%	87,01%	87,19%
Belirli Sözcükler + 12000 sözcük	93,52%	93,34%	92,43%	97,72%
Belirli Sözcükler + 32000 sözcük	93,95%	93,56%	92,87%	92,94%

1.1.2.2 Türkçe İçin Biçimbirimsel Belirsizlik Giderme Kurallarının Öğrenilmesi

[10] daki çalışmada, Türkçe yazıların biçimbirimsel belirsizliğinin giderilmesi için kural tabanlı bir çözüm önerilmiştir. Kurallar karar listeleri öğrenme algoritmasıyla otomatik olarak oluşturulmuştur. Daha önceki bölümlerde anlatılan seyrek veri problemi ve

Türkçe'nin bitişken özelliğinden kaynaklanan problemler, biçimbirimsel çözümleyicinin ürettiği 126 biçimbirimsel özelliğin herbiri için ayrı modeller oluşturulmasıyla çözümlenmiştir. Bir sözcüğün doğru etiketinin bulunması için ilk olarak biçimbirimsel çözümleyici kulllanılarak olası çözümlemeler oluşturulur. Olası çözümlemelerde bulunan herbir özelliğin, doğru çözümlemede olup olmadığına karar listeleri kullanılarak karar verilir. Bu sonuçlar istatistiksel olarak birleştirilerek, sözcüğün olası çözümlemeleri içerisinden doğru olan bulunur.

Bu çalışmayı daha ayrıntılı olarak iki alt başlıkta inceleyebiliriz.

Karar Listeleri

Bilindiği gibi, herbir kuralın bir kalıp ve sınıflandırmadan oluştuğu sıralı kurallar listesine karar listesi adı verilir. Bu çalışmada, ilgili sözcüğün ve çevresindeki sözcüklerin yüzey özelikleri kalıbı oluştururken, ilgili sözcüğün doğru çözümlemesinde bir biçimbirimsel özelliğin olup olmaması ise sınıflandırmayı oluşturur.

Eğer sözcüğün doğru çözümlemesinde +Det özelliği bulunacaksa sınıf sütununda 1, bulunmayacaksa sınıf sütununda 0 olacaktır. Kalıp sütunu ise kuralın eşleşmesi için gerekli olan şartı belirtmektedir. En sondaki kural her hangi bir kalıp içermemektedir ve daha önceki kurallardan hiçbiri uygulanmadıysa, ön değer kuralı olarak uygulanır.Yani karar listeleri en özelden en genele doğru sıralanmıştır. Bir başka deyişle karar listeleri "eğer-öyle-ise" şeklinde yorumlanabilir.

Tablo 1.3: Örnek Karar Listesi

Kural	Sınıf	Kalıp
1	1	W=~çok R1=+DA
2	1	L1=~pek
3	0	W=+AzI
4	0	W=~çok
5	1	-

Tablo da, *+Det* özelliğini tanımlayan örnek bir karar listesi verilmiştir.

Açgözlü Önden Eklemeli Algoritma (AÖA)

Önceki bölümde karar listelerinin nasıl çalıştığından ve ne işe yaradığından bahsettik. Şimdi de karar listelerinin nasıl oluşturulduğunu tartışacağız.

Verilen eğitim örnekleri kümesinden, AÖA kullanılarak karar listeleri oluşturuluyor. İlk olarak liste eğitim kümesinde en sık geçen sınıfa göre hazırlanmış olan ön değer kuralı ile başlar. Daha sonra maksimum fayda sağlanan kural karar listesine eklenir.

Algoritma aşağıda verilmiştir[11]:

```
GPA(data)
1 dlist NIL
2 default-class MOST-COMMON-CLASS(data)
3 rule [if TRUE then default-class]
4 while GAIN(rule; dlist ; data) > 0
5 do dlist prepend(rule; dlist )
6 rule MAX-GAIN-RULE(dlist ; data)
7 return dlist
```

Aday kurallar içerisinde fayda, kuralın listeye eklenmesiyle doğru olarak sınıflandırılan örneklerin sayısındaki artış olarak ele alınır. Aday kurallar listede bulunan herbir kurala yeni nitelik eklenmesiyle bulunur.

Bu çalışmayı alanında eşsiz kılan özellikler şunlardır: (1) istatistiksel ve kural tabanlı teknikleri birleştiren bir öğrenme algoritması ve karar listeleri kullanılması (2) seyrek veri problemine çözüm olarak her bir özelliğin ayrı ayrı ele alınması (3) önceki sözcüklerin etiketlerine bağlı olmaması. Bu çalışmanın başarımı %96 dır ki bu rakam daha önce rapor edilen istatistiksel yöntemlerden daha yüksektir.

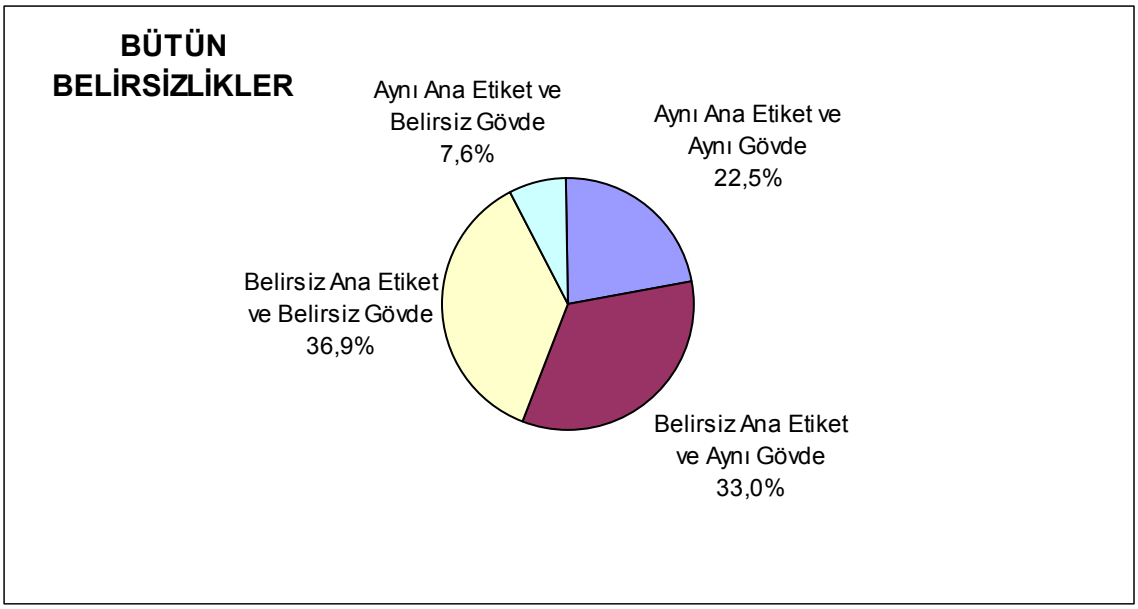
1.2 Tezin Ana Hatları

Bir önceki bölümde tez çalışmamıza yakın olan diğer çalışmalar ayrıntılı biçimde sunulmuştur. Bizim çalışmamızın ana hatları ise aşağıda açıklanmıştır:

Türkçe bir sözcük biçimbirimsel çözümleyiciden geçirildiği zaman, sözcük başına 1,6 çözümleme sonucu ortaya çıkar.

1 milyon sözcükten oluşan derlem üzerinde yaptığımız çalışmalarda şu sonuçlar bulunmuştur. Sözcüklerin %48' inin çözümlemesi tektir. Dolayısıyla bu sözcüklerde biçimbirimsel belirsizlik söz konusu değildir.

Biçim birimsel belirsizlik olan sözcüklerin durumu Şekil 1.1 'de gösterilmiştir.



Şekil 1.1: Biçimbirimsel Belirsizlik Dağılımları

Yukarıda saydığımız temel ilkeler eşiğinde çalışmamızın ilkeleri şöyledir:

İlk aşamada belirsizlik sorunu olmayan sözcükler ayıklandı

Belirsizliği olanlar içinde sıklığı en çok olanlar ele alındı. Belirsizliklerin giderilmesi için Türkçe'nin sözdizimsel ve biçimbirimsel özelliklerine dayalı olarak kural yazıldı.

Yukarıdakilerin dışında kalan sözcükler için kalıplar oluşturulmuştur.

1.3 İçerik

Tezin ikinci kısımında, tezin kapsamıyla sınırlı olmak kaydıyla Türkçe'nin biçimbirimsel yapısı, doğal dil işlemenin özellikleri, Türkçe derlem problemi anlatılmıştır.

Tezin üçüncü kısmı özgün çalışmalarımızın tanıtımına ayrılmıştır. Özetle, bu bölümde biçimbirimsel belirsizlikler sınıflandırılmış, bu sınıflara ilişkin çözüm önerileri getirilmiş, geliştirilmiş kurallar açıklanmıştır. Bu önerilen çözüm yöntemleri denenmiş ve başarımları ölçülmüştür.

Dördüncü bölümde ise, geliştirilmiş olan yazılımlar ve yazılım geliştirme ortamı tanıtılmıştır. Yazılım geliştirme ortamı olarak Komodo, programlama dili olarak ise perl kullanılmıştır.

Beşinci bölüm özgün çalışmaların sınanmasına ve sonuçlarının değerlendirilmesine ayrılmıştır.

Altıncı bölüm sonuç ve önerilere, yedinci bölüm ise kullanılan kaynaklar belirtilmiştir.

Ek A ve Ek B de ise, en sık kullanılan belirsiz sözcüklerin iki farklı derlemde kaç kere geçtiği ve hangi etiketlerle yer aldığının istatistikleri verilmiştir.

2. TEMEL KAVRAMLAR

Bu bölümde tezin konusunu oluşturan temel konular ve kavramlar hakkında bilgi verilecektir. Bundan amaç, tezin bundan sonraki bölümleri için bir altyapı oluşturmaktır.

2.1 Doğal Dil İşleme

Doğal Dil İşleme (DDİ) yapay zeka bilim dalının bir alt dalıdır. DDİ çalışmalarının amacı, bilgisayarların, insanların kullandığı doğal dili anlamasını, insanların bilgisayarla olan ilişkilerinde doğal dillerini kullanabilmesine olanak sağlamaktır. Bu bölümde, ilk olarak, DDİ'nin tanımı, kapsamı ve kaşılaşılan güçlükleri tanıtılacaktır.Daha sonra bu uygulamalar sırasında Türkçe'de karşılaşılan zorluklara değinilecektir.

Doğal Dil İşleme temel olarak 5 düzeyden oluşmaktadır[13]:

- Sesbilim (sözcüklerin seslendirilişi ile ilgilidir.)
- Biçimbirim (sözcüklerin yapısı ile ilgilidir)
- Sözdizim (sözcüklerin tümce içindeki sıralanışını inceler)
- Anlambilim (sözcüklerin anlamıyla ilgilenir)
- Söylem (dilin kullanımı)

Sesbilim

Sesbilim bir dilde seslerin nasıl kullanıldığıyla ilgili çalışmaları kapsamaktadır.

Her dilde, onu diğer dillerden ayıran bir ses abecesi vardır. Bunların herbirine sesbirim adı verilir. Türkçe abecesi 29 harften oluşur ve bilinen diller arasında, bir harfe karşı bir ses düştüğü varsayılır. Ancak, bu alanda çalışanların ortaya koyduğuna göre, Türkçenin sesbirim sayısı 45 tir. Bunlardan 15 tanesi sesli, 27 tanesi ise sessizdir [12].

Biçimbirim

Biçimbirim sözcüklerin yapısıyla ile ilgilenir. Türkçe için söcüklerin türetilmesi ve ekler çok önem taşır. Her dilde iki farklı şekilde sözcük oluşturulabilir. Bunlardan biri çekim,

diğeri ise türetme yöntemidir. Çekim yoluyla sözcük oluşturulurken bir sözcüğün farklı şekilleri kullanılır. *çiçek* sözcüksine çekim eki eklenmesiyle *çiçekler* sözcüğünün ortaya çıkması bu yolla oluşturulan sözcüklere örnek olarak verilebilir. Türetme ise var olan eski sözcüklere yapım ekleri eklenmesi yoluyla yeni sözcük oluşturma yöntemidir. Sözgelimi, *göz* sözcüğünden *gözlük* sözcüğünün türetilmesi buna bir örnektir. Özellikle bu tezi ilgilendiren düzey biçimbirim düzeyidir.

Bir sözcüğün bütün biçimbirimsel formlarının bulunmasına biçimbirimsel çözümleme denir. Örneğin, *aile* sözcüğü biçimbirimsel çözümleyiciden geçirilirse şu sonuç karşımıza çıkar:

aile **(** ***aile*****+**Noun+ A3sg+ Pnon+ Nom)

Noun: İsim

*A3sg :*3.tekil şahıs ekini ifade ediyor

Pnon: İyelike eki alamdığını gösteriyor

Nom: İsimin yalın hali olduğunu belirtiyor

Bu çözümleme sonucunda, aile sözcüğünün 3.tekil şahış eki almış bir isim olduğu ve yalın halde bulunduğu bilgilerine ulaşabiliyoruz.

Sözdizim

Sözdizim bir tümcenin içindeki sözcüklerin dizilişi ile ilgilenen düzeydir. İnsanlar konuşurken ya da yazarken var olan sözcükleri belli bir sırada yerleştirerek yeni tümceler oluştururlar. Bu nedenle sözdizimsel düzey, sesbilim ve biçimbirim düzeylerinin üstünde yer alır.

Anlambilim

Anlambilim düzeyi dilin anlamıyla ilgilenir, böylelikle gerçek dünyayla dil arasında bağlantı kurar. Bu düzeyde yapılan çalışmalar daha önce anlatılan düzeylere göre çok azdır.

Söylem

Söylem düzeyi dilin bağlamda kullanımıdır. Anlambilim ve söylem düzeyleri arasındaki fark çok küçüktür. Bazı araştırmacılar bu iki düzeyi birlikte ele alır.

2.2 Biçimbirimsel Belirsizlik

Her sözcüğün biçimbirimsel çözümlenmesi sonunda tek bir sonuç üretilmeyebilir. Bu durumda cümle içinde geçen sözcüğün hangi çözümleme sonucunun doğru olduğuna karar verilmesi gerekir. Bu duruma biçimbirimsel belirsizlik denir.Aşağıda *koyun* sözcüğünün biçimbirimsel çözümleme sonuçları gösterilmiştir.

–koy (koy+Noun+ A3sg+ P2sg+ Nom) (Kalamış koyu)

–koy (koy+Noun+ A3sg+ Pnon+ Gen) (Kalamış koyunun)

–koy (koy+Verb+ Pos+ Imp+ A2pl) (kitapları masanın üzerine koyun)

–koyu (koyu+Adj^DB+ Noun+ Zero+ A3sg+ P2sg+ Nom) (koyu rengin)

–koyun (koyun+Noun+ A3sg+ Pnon+ Nom) (hayvan olan koyun)

2.3 Türkçe'nin Biçimbirimsel Yapısı

Türkçe Ural – Altay dil ailesine mensup bitişken bir dildir. Bu bölümde Türkçe'nin biçimbirimsel özellikleri incelenmiş, dil yapısından kaynaklanan problemler ele alınmıştır.

Türkçe'nin biçimbirimsel yapısı üzerinde uygulama geliştirmek, bitişken olmayan diğer dillere göre daha zordur. Bunun iki nedeni vardır. Bunlardan birincisi Türkçe'deki sözcüklerin sayısının teorik olarak sonsuz olmasıdır. Çünkü bir sözcüğün alabileceği yapım ve çekim eklerinin sayısı sınırsızdır.

Örneğin, Tablo 2.1'de gösterilen sözcükleri İngilizce karşılıklarını yazmayı denersek, Türkçe'de tek bir sözcükle ifade edebildiğimiz anlamı, İngilizce'de ancak bir cümleyle ifade edebiliriz.

Tablo 2.1: Örnek Fiil Çekimi

görüyorum	gördük	görmeyince
görüyorsun	gördünüz	gördükçe
görüyor	gördüler	görmeliyiz
görüyoruz	görmüşüm	görmeden
görüyorsunuz	görmüşsün	görünce
görüyorlar	görmüş	gördürmek
gördüm	görmüşüz	gördürtmek
gördün	görmüşsünüz	gördürttürmek
gördü	görmüşler	

Türkçe sondan eklemeli bir dildir ve eklerin çeşidi de diğer dillere göre fazladır. Ön ek açısından çok kısıtlı olmasına karşın, son ekler konusunda olağanüstü zengindir. Bu eklerin 300 civarında bulunduğu ve bunların yarısının yoğun bir şekilde kullanıldığı bilinmektedir. Fiilden fiiil yapan 10, fiilden isim yapan 54, isimden fiil yapan 12, isimden isim yapan 80 ek yoğun biçimde kullanılmaktadır [12]. Bu eklerin sayısı fazla olduğu kadar, eklerden türetilebilcek sözcük sayısı da fazladır.

Örneğin, –ecek –acak eki ile oluturulabilecek sözcükler Tablo 2.2'de verilmiştir.

Tablo 2.2: -ecek/-acak Ekiyle Oluşan Sözcükler

acınacak	çıkacak	giyecek	salacak
acınılacak	çizecek	içecek	sıkacak
açacak	delecek	kıracak	silecek
alacak	diyecek	koyacak	tutacak
alacak verecel	doğacak	olacak	verecek
çalacak	ekecek	olmayacak	yakacak
çekecek	gelecek	oturacak	yiyecek

-

Türkçe'nin tüm bu özelliklerden dolayı biçimbirimsel etiketlerin sayısı da sınırsızdır. Türkçe'nin bu özelliği yapılan istatistiksel yöntemlerde seyrek veri problemine yol açmaktadır. Bu da biçimbirimsel belirsizlik giderme ve etiketleme işlemlerini Türkçe için bir kat daha zorlaştırmaktadır.

Türkçe'de sözcüklerin dizilişi genellikle özne – nesne – yüklem şeklinde sıralanır. Ancak bu sıranın dışında isim tamlamaları cümlelerde farklı dizilişlerde yer alabilir. Sadece fiilin yeri sabittir ve cümlenin en sonunda bulunur.

Türkçe'nin sabit bir sözcük diziliş sırası olmadığından, n-gram Hidden Markov yöntemlerinin İngilizce'deki kadar başarısı Türkçe'de yakalanamamıştır.

2.4 Derlem

Tezde, eğitim ve çeşitli istatistiklerin çıkartılması amacıyla daha önceden elle doğruları işaretlenmiş bir derleme ihtiyaç duyulmuştur. Ancak ne yazık ki ,Türkçe için elle işaretlenmiş büyük ölçekli bir derlem bulunmamaktadır. Türkçe için var olan ODTÜ-SABANCI Ağaç Yapılı Derlemi [9] tümüyle elle işaretlenmiş olmasına rağmen küçük ölçekli bir derlem olması itibariyle istatistik çıkarmak için çok uygun değildir. Daha büyük ölçekli olan derlem ise yarı otomatik işaretlendiğinden hata oranı Ağaç Yapılı derleme göre çok fazladır.

Biz de istatistik çıkarmak için büyük ölçekli derlemi, kurallarımızın doğruluğunu sınamak için ise ağaç yapılı derlemi kullandık.

3. TÜRKÇE İÇİN KURAL TABANLI BİÇİMBİRİMSEL BELİRSİZLİK GİDERİCİ

Türkçe kurallı bir dil olduğundan, biçimbirimsel berlirsizliklerin giderilmesi için de Türkçe'nin dilbilgisine uygun kurallar yazılabilir. Bu tezde, kuralları elle yazılmış kural tabanlı biçimbirimsel belirsizlik giderici sunulmuştur.

Bu bölümde, Türkçe'deki biçimbirimsel belirsizliklerden bahsedilmiş ve bu belirsizliklerin giderilmesi için çözümler geliştirilmiştir.

3.1 Türkçe'deki Biçimbirimsel Belirsizlikler

Tezde, analizlerin yapılması için kullandığımız derlem noktalama işaretleri hariç 728662 sözcük içermektedir. Bu derlemdeki sözcükleri biçimbirimsel çözümleyiciden geçirdiğimizde görülmüştür ki, sözcüklerin 382725 tanesi için birden fazla biçimbirimsel çözümleme ortaya çıkıyor.Yani toplam derlemin %52si belirsiz sözcüklerden oluşuyor.

Bu bölümde bu belirsizliklerin kaynakları oluşmaktadır.

3.1.1 Gövde Belirsizlikleri

Derlemdeki %52 belirsizlikler incelendiğinde bunların %44,5' inde gövde belirsizliği olduğu belirlenmiştir (bknz Şekil 3.1). Derlemde bulunan. Biçimbirimsel çözümleyiciden çıkan olası çözümlemeleri içerisinde, 170460 sözcüğün gövde belirsizliği olduğu görülmüştür.

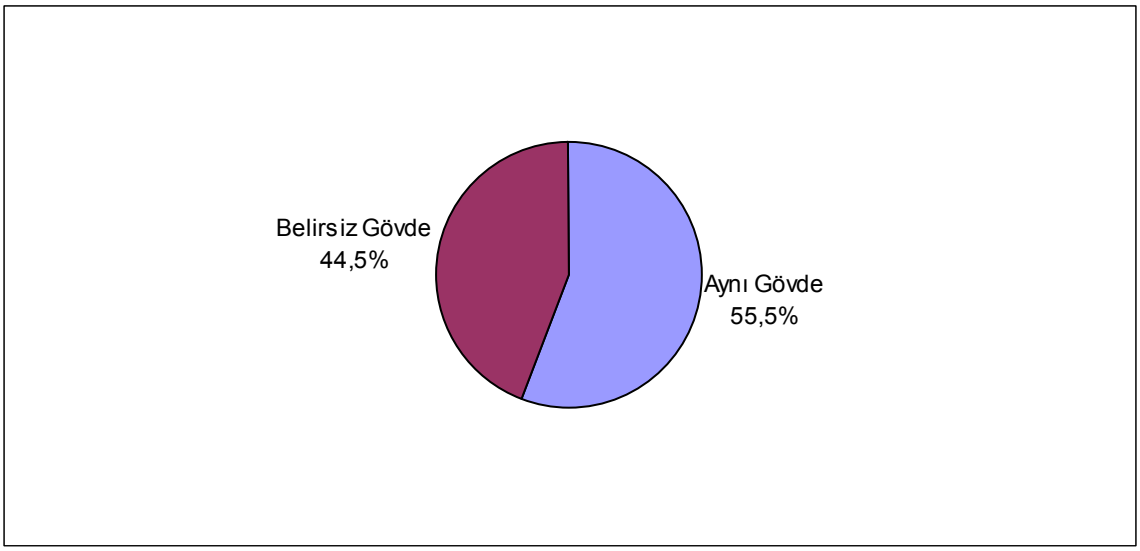
Örneğin, "araba" sözcüğünü biçimbirimsel çözümleyiciden geçirdiğimizde iki sonuç ortaya çıkar:

araba sözcüğünün biçimbirimsel çözümleyici çıktısı:

arap (***arap***+Noun+ A3sg+ Pnon+ Dat)

araba (***araba***+Noun+ A3sg+ Pnon+ Nom)

Burada gövde binek olan ***araba*** veya ulusu belirten ***arap*** sözcüğü olabilir. Bu durumda ***araba*** sözcüğünün biçimbirmsel çözümlemeleri arasında gövde belirsizliği vardır denir.



Şekil 3.1: Gövde Belirsilikleri

3.1.2 Ana Etiket Belirsizlikleri

Derlemdeki %52 belirsizlikler incelendiğinde bunların %69,8 inde ana etiket belirsizliği olduğu belirlenmiştir (bknz Şekil 3.2). Derlemde bulunan. Biçimbirimsel çözümleyiciden çıkan olası çözümlemeleri içerisinde, 267266 sözcüğünana etiket belirsizliği olduğu görülmüştür.

Örneğin, "kalem" sözcüğünün olası biçimbirimsel çözümlerimelerinin içerisinde ana etiket belirsizliği yoktur. Çünkü her iki çözümlemede de ana etiket isimdir.

kalem sözcüğünün biçimbirimsel çözümleyici çıktısı:

kale (***kale***+Noun+ A3sg+ P1sg+ Acc)

kalem (***kalem***+Noun+ A3sg+ Pnon+ Nom)

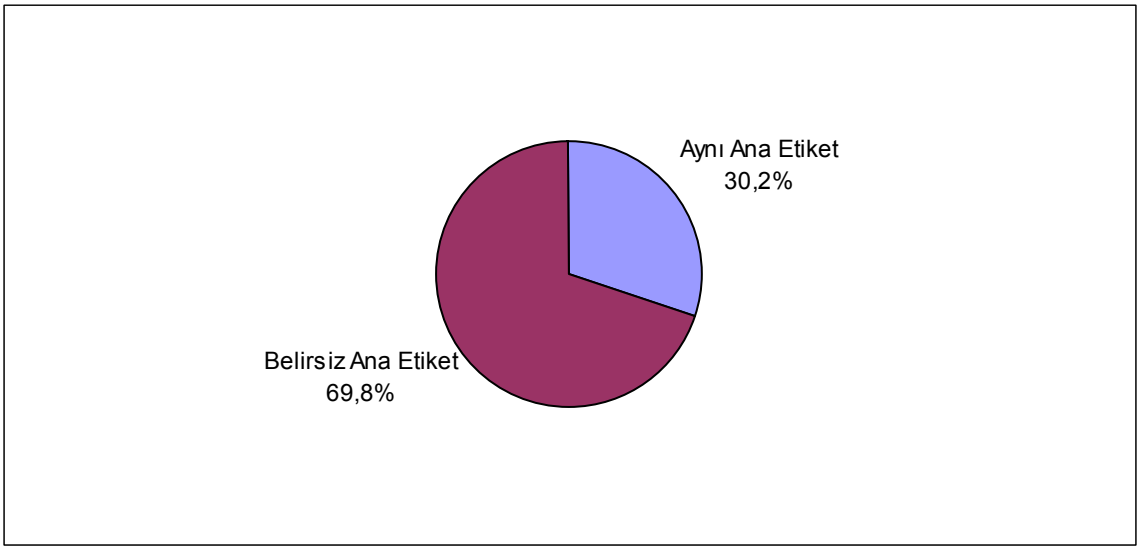
Oysa, "yüz" sözcüğünün olası biçimbirimsel çözümlerimelerinin içerisinde ana etiket belirsizliği vardır. İlk çözümlemede sayı sıfatı, ikinci de isim, üçüncüde ise fiil ana etiketlerine sahiptir.

yüz sözcüğünün biçimbirimsel çözümleyici çıktısı:

yüz (***yüz***+Num+ Card)

yüz (***yüz***+Noun+ A3sg+ Pnon+ Nom)

yüz (***yüz***+Verb+ Pos+ Imp+ A2sg)



Şekil 3.2: Ana Etiket Belirsizlikleri

3.1.3 Çekim Grubu Belirsizlikleri

Derlemdeki %52 belirsizlikler incelendiğinde bunların %22,5 inde sadece çekim grubu belirsizliği olduğu belirlenmiştir (bknz Şekil 3.3). Derlemde bulunan. Biçimbirimsel çözümleyiciden çıkan olası çözümlemeleri içerisinde, 86244 sözcüğünçekim grubu belirsizliği olduğu görülmüştür.

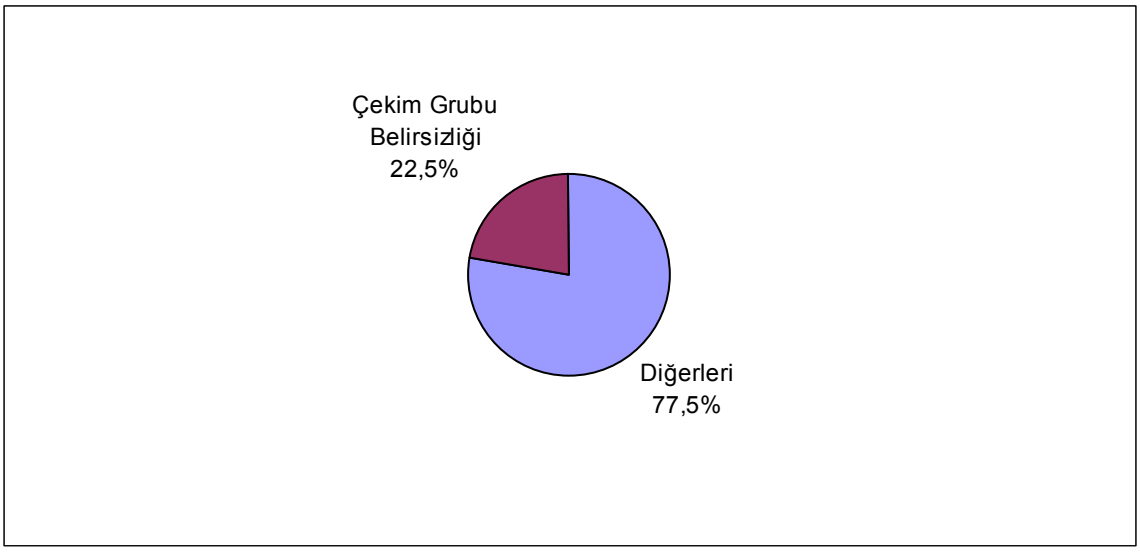
Örneğin, "çocuğu" sözcüğünü biçimbirimsel çözümleyiciden geçirdiğimizde iki sonuç ortaya çıkar.

çocuğu sözcüğünün biçimbirimsel çözümleyici çıktısı:

çocuğu (***çocuk***+Noun+ A3sg+ P3sg+ Nom)

çocuğu(***çocuk***+Noun+ A3sg+ Pnon+ Acc)

Görüldüğü gibi bu örnekte hem gövdeler aynı ***(çocuk),*** hem de ana etiketler ***(isim)*** aynıdır. Belirsizliği oluşturan çekim gruplarıdır (***P3sg+ Nom/ Pnon+ Acc).***



Şekil 3.3: Çekim Grubu Belirsizlikleri:

3.1.4 Diğer Belirsizlik Analizleri

Derlemdeki %52 belirsizlikler incelendiğinde bunların %7,6 sında ana etiketlerin aynı fakat gövdelerin belirsizlik içerdiği gözlemlenmiştir.

Derlemdeki %52 belirsizlikler incelendiğinde bunların %36,9 sında ana etiketlerin de gövdelerin de belirsizlik içerdiği gözlemlenmiştir.

Derlemdeki %52 belirsizlikler incelendiğinde bunların %14,4 sında gövdelerin aynı fakat ana etiketlerin belirsizlik içerdiği gözlemlenmiştir.

3.1.5 Kalıplar

Sadece çekim gruplarından kaynaklanan belirsizlikler incelendiğinde, bunların içinde belli kalıpların sürekli tekrarlandığı tespit edilmiştir. Örneğin "içinde" kelimesi için biçimbirimsel çözümleyiciden geçirildiğinde ortaya çıkan sonuçlar arasındaki farkın sadece çekim gruplarında olduğu gözlenmiştir.

içinde sözcüğünün biçimbirimsel çözümleyici çıktısı:

iç (***iç***+Noun+ A3sg+ P3sg+ Loc)

iç (***iç***+Noun+ A3sg+ P2sg+ Loc)

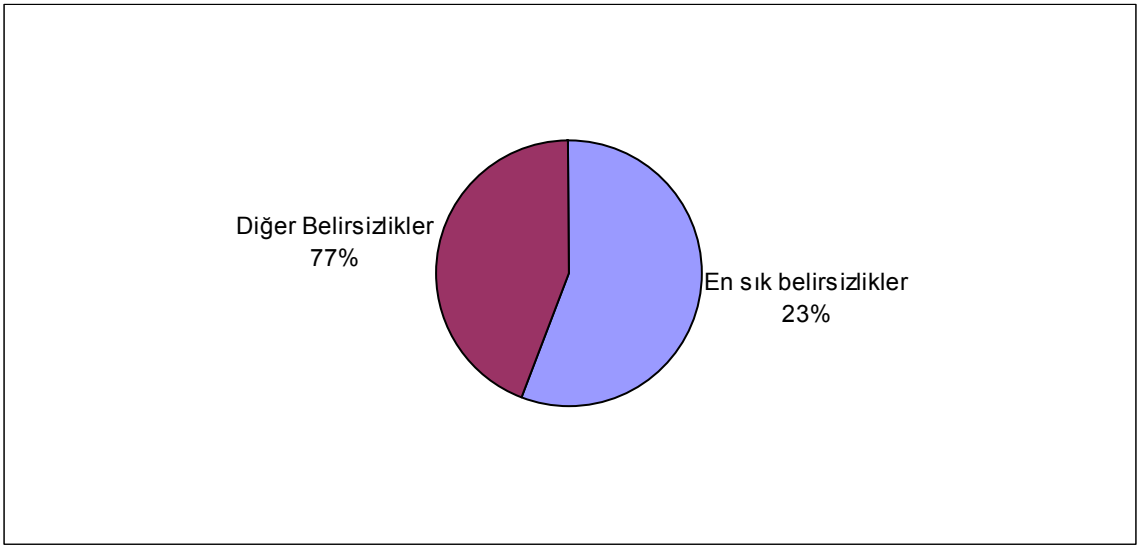
Bu belirsizlik P3sg (3.tekil şahıs, iyelik eki) ve P2sg (2.tekil şahıs, iyelik eki) eklerinden kaynaklanmaktadır. Derlemimizi tradığımızda sadece bu iki ekten kaynaklanan belirsizliklerin toplam belirsizliğin %7,29 'u olduğunu gördük. Başka bir deyişle sadece

bu iki ek arasındaki belirsizliğin çözülmesiyle, toplam belirsizliğin büyük bir kısmı çözümlenmiş olacaktır.

3.2 Türkçe'de En Sık Kullanılan Belirsiz Sözcükler

Büyük ölçekli derlemi taradığımızda görülmüştür ki, Türkçe'de biçimbirimsel belirsizliği artıran sözcükler bir grup sözcükler. Bunlar çoğunlukla bağlaç veya edat görevi gören, ek almadan kullanılan sözcüklerdir (bknz: EK A) Bu sözcükler toplam belirsizliğin %23'ünü oluşturmaktadır.

Eğer bu kısımlar için kural yazılırsa büyük bir kısmın belirsizliği giderilmiş olacaktır.



Şekil 3.4: En Sık Kullanılan Belirsizlik Sözcükler

En sık kullanılan belirsizliz sözcüklerin, belirsizliğinin giderilmesi için aşağıdaki kurallar yazılmıştır:

bir Bir - - - bir+Det

</R>

bu Bu NT 0 Noun bu+Det

bu Bu NT 0 Conj bu+Pron+Demons+A3sg+Pnon+Nom

bu Bu NT 0 Punc bu+Pron+Demons+A3sg+Pnon+Nom

bu Bu NT 0 Verb bu+Pron+Demons+A3sg+Pnon+Nom

bu Bu - - - bu+Det

</R>

de De - - - de+Conj

</R>

için İçin - - - için+Postp+PCNom

</R>

çok Çok NT 2 Verb çok+Adverb

çok Çok NT 2 Adverb çok+Adverb

çok Çok NT 2 Noun çok+Adj

çok Çok NT 2 Num çok+Adj

çok Çok NT 2 Adj çok+Adverb

çok Çok PT 2 Abl çok+Postp+PCAbl

çok Çok - - - çok+Adverb

</R>

en En - - - en+Adverb

</R>

sonra Sonra PT 0 Abl sonra+Postp+PCAbl

sonra Sonra - - - sonra+Adverb

</R>

ama Ama - - - ama+Conj

</R>

ne Ne - - - ne+Pron+QuesP+A3sg+Pnon+Nom

</R>

kadar Kadar PT 0 Dat kadar+Postp+PCDat

kadar Kadar - - - kadar+Noun+A3sg+Pnon+Nom

</R>

var Var - - - var+Adj^DB+Verb+Zero+A3sg

</R>

ise İse - - - ise+Conj

</R>

değil Değil NT 0 Noun değil+Conj

değil Değil NT 0 Punc değil+Verb+Pres+A3sg

değil Değil - - - değil+Verb+Pres+A3sg

</R>

yeni Yeni - - - yeni+Adj

</R>

ancak Ancak NT 2 Verb ancak+Adverb

ancak Ancak NT 2 Adverb ancak+Adverb

ancak Ancak - - - ancak+Conj

</R>

diye Diye - - - diye+Postp+PCNom

</R>

ilk İlk NT 2 Verb ilk+Adverb

ilk İlk NT 2 Adverb ilk+Adverb

ilk İlk - - - ilk+Adj

</R>

son Son NT 0 Punc son+Noun+A3sg+Pnon+Nom

son Son NT 2 Verb son+Noun+A3sg+Pnon+Nom

son Son - - - son+Adj

</R>

yıl Yıl - - - yıl+Noun+A3sg+Pnon+Nom

</R>

yüzde Yüzde - - - yüz+Num+Card^DB+Noun+Zero+A3sg+Pnon+Loc

</R>

önce Önce PT 0 Abl önce+Postp+PCAbl

önce Önce - - - önce+Adverb

</R>

göre Göre - - - göre+Postp+PCDat

</R>

o O NT 2 Verb o+Pron+Demons+A3sg+Pnon+Nom

o O - - - o+Det

</R>

yok Yok - - - yok+Adj^DB+Verb+Zero+A3sg

</R>

karşı Karşı PT 2 Dat karşı+Postp+PCDat

karşı Karşı - - - karşı+Noun+A3sg+Pnon+Nom

</R>

ya Ya - - - ya+Conj

</R>

ilgili İlgili - - - ilgili+Noun+A3sg+Pnon+Nom

</R>

böyle Böyle NT 2 Verb böyle+Adverb

böyle Böyle NT 2 Adverb böyle+Adverb

böyle Böyle - - - böyle+Adj

</R>

bile Bile - - - bile+Adverb

</R>

ben Ben - - - ben+Pron+PersP+A1sg+Pnon+Nom

</R>

yer Yer NT 2 Verb yer+Verb+Pos+Imp+A2sg

</R>

başka Başka PT 0 Abl başka+Potsp+PCAbl

başka Başka - - - başka+Adj

</R>

bin Bin - - - bin+Num+Card

</R>

bazı Bazı NT 2 Verb bazı+Det

bazı Bazı - - - bazı+Adj

</R>

hiç Hiç - - - hiç+Adverb

</R>

hem Hem NT 2 Verb hem+Adverb

hem Hem - - - hem+Conj

</R>

şöyle Şöyle NT 2 Verb şöyle+Adverb

şöyle Şöyle NT 2 Adverb şöyle+Adverb

şöyle Şöyle - - - şöyle+Adj

</R>

mi Mi - - - mi+Ques+Pres+A3sg

</R>

başka Başka PT 0 Abl başka+Potsp+PCAbl

başka Başka - - - başka+Adj

</R>

gelen Gelen - - - gel+Verb+Pos^DB+Adj+PresPart

</R>

şimdi Şimdi - - - şimdi+Adverb

</R>

nasıl Nasıl NT 2 Verb nasıl+Adverb

nasıl Nasıl NT 2 Adverb nasıl+Adverb

nasıl Nasıl - - - nasıl+Adj

</R>

sadece Sadece - - - sadece+Adverb

</R>

birlikte Birlikte PT 2 Ins birlikte+Postp+PCIns

birlikte Birlikte - - - birlikte+Adverb

</R>

eski Eski - - - eski+Adj

</R>

artık Artık NT 2 Verb artık+Adverb

artık Artık NT 2 Adverb artık+Adverb

artık Artık - - - artık+Adj

</R>

şu Şu NT 2 Verb şu+Pron+Demons+A3sg+Pnon+Nom

şu Şu - - - şu+Det

</R>

doğru Doğru PT 2 Dat doğru+Postp+PCDat

doğru Doğru - - - doğru+Adj

</R>

devam Devam - - - devam+Adj

</R>

tam Tam NT 2 Verb tam+Adverb

tam Tam NT 2 Adverb tam+Adverb

tam Tam - - - tam+Adj

</R>

neden Neden NT 2 Verb neden+Adverb

neden Neden - - - ne+Pron+QuesP+A3sg+Pnon+Abl

</R>

fazla Fazla NT 2 Verb fazla+Adverb

fazla Fazla NT 2 Adverb fazla+Adverb

fazla Fazla PT 2 Abl fazla+Postp+PCAbl

fazla Fazla - - - fazla+Adj

</R>

buna Buna - - - bu+Pron+Demons+A3sg+Pnon+Dat

</R>

ayrıca Ayrıca - - - ayrıca+Adverb

</R>

altında Altında - - - alt+Noun+A3sg+P3sg+Loc

</R>

eğer Eğer - - - eğer+Conj

</R>

uzun Uzun - - - uzun+Adj

</R>

dış Dış - - - dış+Adj

</R>

geri geri - - - geri+Adverb

</R>

onun Onun - - - o+Pron+PersP+A3sg+Pnon+Gen

</R>

az Az NT 2 Verb az+Adverb

az Az NT 2 Adverb az+Adverb

az Az PT 2 Abl az+Postp+PCAbl

az Az - - - az+Adj

</R>

açık Açık NT 2 Verb açık+Adverb

açık Açık NT 2 Adverb açık+Adverb

açık Açık - - - açık+Adj

</R>

iç İç - - - iç+Noun+A3sg+Pnon+Nom

</R>

kadın Kadın - - - kadın+Noun+A3sg+Pnon+Nom

</R>

dışında Dışında - - - dış+Noun+A3sg+P3sg+Loc

</R>

yana Yana PT 2 Abl yana+Postp+PCAbl

yana Yana - - - yan+Noun+A3sg+Pnon+Dat

</R>

hatta Hatta - - - hatta+Conj

</R>

dek Dek PT 2 Dat dek+Postp+PCDat

dek Dek - - - dek+Noun+A3sg+Pnon+Nom

</R>

gerek Gerek - - - gerek+Verb+Pos+Imp+A2sg

</R>

orta Orta - - - orta+Adj

</R>

sık Sık NT 2 Verb sık+Adverb

sık Sık NT 2 Adverb sık+Adverb

sık Sık - - - sık+Adj

</R>

yakın Yakın - - - yakın+Adj

</R>

kısa Kısa - - - kısa+Adj

</R>

sonunda Sonunda - - - sonunda+Adverb

</R>

onu Onu - - - o+Pron+PersP+A3sg+Pnon+Acc

</R>

belli Belli - - - belli+Adj

</R>

biraz Biraz NT 2 Verb biraz+Adverb

biraz Biraz NT 2 Adverb biraz+Adverb

biraz Biraz - - - biraz+Adj

</R>

beni Beni - - - ben+Pron+PersP+A1sg+Pnon+Acc

</R>

ileri İleri - - - ileri+Adj

</R>

hala Hala - - - hala+Adverb

</R>

öyle Öyle NT 2 Verb öyle+Adverb

öyle Öyle NT 2 Adverb öyle+Adverb

öyle Öyle - - - öyle+Adj

</R>

hale Hale - - - hal+Noun+A3sg+Pnon+Dat

</R>

türlü Türlü - - - türlü+Adj

</R>

böylece Böylece - - - böylece+Adverb

</R>

onlar Onlar - - - o+Pron+PersP+A3pl+Pnon+Nom

</R>

üst Üst - - - üst+Adj

</R>

bizi Bizi - - - biz+Pron+PersP+A1pl+Pnon+Acc

</R>

beri Beri PT 2 Abl beri+Postp+PCAbl

beri Beri - - - beri+Adverb

</R>

oysa Oysa NT 2 Verb oysa+Adverb

oysa Oysa NT 2 Adverb oysa+Adverb

oysa Oysa - - - oysa+Conj

</R>

boyunca Boyunca - - - boyunca+Postp+PCNom

</R>

aslında Aslında - - - aslında+Adverb

</R>

amacıyla Amacıyla - - - amaç+Noun+A3sg+P3sg+Ins

</R>

güney Güney - - - güney+Adj

</R>

ona Ona - - - o+Pron+PersP+A3sg+Pnon+Dat

</R>

işte İşte - - - işte+Adverb

</R>

başta Başta - - - başta+Adverb

</R>

bütün Bütün - - - bütün+Adj

</R>

onların Onların - - - o+Pron+PersP+A3pl+Pnon+Gen

</R>

yoksa Yoksa - - - yok+Adj^DB+Verb+Zero+Cond+A3sg

</R>

karşın Karşın PT 2 Dat karşın+Postp+PCDat

karşın Karşın - - - karşı+Noun+A3sg+P2sg+Nom

</R>

yalnız Yalnız NT 2 Verb yalnız+Adverb

yalnız Yalnız NT 2 Adverb yalnız+Adverb

yalnız Yalnız - - - yalnız+Adj

</R>

4. GEREKSİNİMLER VE PROGRAMLAMA ORTAMI

Bu tezdeki uyguama yazılım tamamen Perl dilinde yazıldı. Bu yazılımı makinanızda çalıştırmak için perl ortamı kurulmalıdır. Çalıştırmak için her hangi bir ekstra donanıma ihtiyaç duyulmamaktadır. İhtiyaç olan tek şey üzerinde perl çalıştırılabilecek bir bilgisayardır.

Günümüzde mevcut bulunan bütün programlama dillerinden , doğal dil işleme çalışmaları için en uygun olan dillerden birisi Perl dilidir. Bunun nedenleri arasında şunlar sayılabilir:

Doğal dil işleme uygulamalarında çok önemli olan düzenli ifadelerin yazımı ve programa entegre edilmesi perl diliyle çok kolaydır.

Doğal dil işleme uygulamalarına yönelik bir çok hazır doğal dil modülleri vardır.

Bütün platformlarda çalışabilir ve diğer platformlara taşınması kolaydır.

Birçok Unix aracı Perl diline entegre edilebilir.

Perl benzerlerinden daha hızlı çalışır.

Lisp dili bu avantajların sadece birkaçına sahiptir. Pascal ve C gibi geleneksel diller de bu avantajların bir çoğundan yoksundur. Elbette başka dillerle de doğal dil işleme uygulamaları yapılabilir ancak perl ile bu çok daha kolaydır.

5. SINAMA

Sınama, yazılım üretme aşamasında çok önemli bir kavramdır.Yanlışları gidermek için program defalarca sınanmalıdır.Bu sınamaların amacı kaynaktaki problemleri bulmaktır.

Bu tezde, her parça tamamlandıktan sonra sınama yapılmıştır.Bu yolla hatalar daha kolay tespit edilip, daha çabuk düzeltilmiştir. Ayrıca sisteme yeni bir parça eklemeden eskinin eksizsiz çalıştığından emin olunmuştur.

Sınamalar olabililecek tüm koşullar göz önünde bulunundurularak yapılmıştır. Sisteme yeni bir kural eklendiğinde onun getirdiği artılar kadar eksiler de hesaba katılarak başarımı değerlendirilmiştir. her kural bazında başarımlar verilmiştir. Toplam başarım %82,6dır.

Örneğin, ***bu*** sözcüğü için yazılan kural 554 kere uygulanmış, bunlardan 57 tanesi hatalı etiketlemiştir. ***bu*** sözcüğü için başarım %89,71dir.

Sözcük	KUS	Hata	Başarım
bir	1214	150	87,64%
bu	554	57	89,71%
de	433	1	99,77%
için	229	0	100%
en	108	0	100%
sonra	140	39	72,14%
ama	192	0	100%
ne	170	46	72,94%
kadar	148	66	55,41%
var	108	52	51,85%
ise	66	17	74,24%
değil	101	38	62,38%
yeni	76	4	94,74%
ancak	53	12	77,36%
diye	172	3	98,26%
ilk	60	12	80%
son	57	10	82,46%
yıl	43	2	95,35%
yüzde	23	2	91,30%
önce	94	45	52,13%
göre	51	0	100%
o	201	74	63,18%

yok	57	36	36,84%
karşı	20	6	70%
ya	57	10	82,46%
ilgili	38	18	52,63%
böyle	67	30	55,22%
bile	51	2	96,08%
ben	187	1	99,47%
yer	14	14	0%
başka	57	11	80,70%
bin	25	7	72%
bazı	28	1	96,43%
hiç	75	3	96%
hem	45	7	84,44%
şöyle	28	14	50%
mi	66	0	100%
başka	57	11	80,70%
gelen	28	0	100%
şimdi	58	0	100%
nasıl	57	29	49,12%
sadece	22	0	100%
birlikte	36	17	52,78%
eski	34	1	97,06%
artık	53	46	13,21%
şu	36	4	88,89%
doğru	56	5	91,07%
devam	0	0	0%
tam	31	13	58,06%
neden	24	11	54,17%
fazla	13	1	92,31%
buna	22	0	100%
ayrıca	8	0	100%
altında	12	0	100%
eğer	13	1	92,31%
uzun	35	0	100%
dış	6	0	100%
geri	16	9	43,75%
onun	65	11	83,08%
az	22	8	63,64%
açık	13	2	84,62%
iç	11	7	36,36%
kadın	46	0	100%
dışında	10	0	100%
yana	12	2	83,33%
hatta	21	0	100%
dek	6	1	83,33%
gerek	13	6	53,85%
orta	7	1	85,71%
sık	5	3	40%

yakın	9	3	66,67%
kısa	13	1	92,31%
sonunda	18	2	88,89%
onu	70	11	84,29%
belli	16	1	93,75%
biraz	52	22	57,69%
beni	91	2	97,80%
ileri	3	2	33,33%
hala	14	2	85,71%
öyle	42	18	57,14%
hale	6	2	66,67%
türlü	13	7	46,15%
böylece	13	2	84,62%
onlar	16	4	75%
üst	7	1	85,71%
bizi	14	0	100%
beri	12	1	91,67%
oysa	18	5	72,22%
boyunca	17	1	94,12%
aslında	26	0	100%
amacıyla	4	0	100%
güney	1	1	0%
ona	60	9	85%
işte	34	4	88,24%
başta	3	1	66,67%
bütün	14	0	100%
onların	21	4	80,95%
yoksa	15	5	66,67%
karşın	6	0	100%
yalnız	10	2	80%
Toplamda %83,43			

6. SONUÇLAR ve TARTIŞMA

Türkçe'deki biçimbirimsel belirsizlikler incelenmiştir. Türkçe'nin bitişken bir dil olmasından dolayı, sözcüklerin %52 sinin birden fazla çözümünün olduğu, her bir sözcük başına 1,8 çözümleme düştüğü saptanmıştır.

Türkçe'deki biçimbirimsel belirsizlikler sınıflandırılmıştır.

Öncelikle, en sık kullanılan belirsiz sözcüklerin belirsizliğinin giderilmesi için kurallar yazılmıştır. Geriye kalan sözcükler için kalıplar bulunmuş ve bu kalıplara göre kurallar yazılmıştır. Kısacası sınıflara ayrılan belirsizliklere çözüm önerilerinde bulunulmuştur.

Sonunda elde edilen çözüm sınanmıştır. Buna göre yazdığımız kurallarla en sık kullanılan belirsiz sözcüklerin belirsizliği %82,6 oranında giderilmiştir.

7. KAYNAKLAR

[1] **Brill, E.,** 1995. Transformation – Based Error – Driven Learning and Natural Language Processing : A Case Study in Part of Speech Tagging, *In Computational Lingusistics,* 21(4): 543-565.

[2] **Brill, E.,** 1992. A Simple Rule – Based Part of Speech Tagger, *Proceedings of* the *DARPA Speech and Natural Language Workshop*, Morgan Kauffman, San Mateo, California, 112-116.

[3] **Megyesi, B.,** 1999. Improving Brill's PoS Tagger for an Agglutinative Language**,** *Proceedings of the Joint Sigdat Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, University of Maryland, USA, June 1999, 275-284.

[4] **Megyesi, B.,** 1999. Brill's PoS Tagger with Extended Lexical Templates for Hungarian, *Proceedings of the Workshop (W01) on Machine Learning in Human Language Technology*, ACAI'99, Greece, July 1999, 22-28.

[5] **Hakkani – Tür D. Z., Oflazer, K. and Tür G.,** 2002. Statistical Morphological Disambiguation for Agglutinative Languages, *Computers and the Humanities*, 381-410.

[6] **Oflazer K.,** 1994. Two-Level Description of Turkish Morphology, *Literary and Linguistic Computing*, 137-148.

[7] **Oflazer K. and Kuruöz İ.,** 1994. Tagging and Morphological Disambiguation of Turkish Text, *Proceedings of the 4th Applied Natural Language Processsing Conference, 144-149*

[8] **Oflazer K. and Tür G.,** 1997. Morphological Disambiguation by Voting Constraints, *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics,*Madrid, Spain.

[9] **Oflazer, K., Hakkani – Tür D. Z. and Tür G.,** 1999. Design for a Turkish Treebank. *Proceedings of the Workshop on Linguistically Interpreted Corpora*, Madrid, Spain, June 1999.

[10] **Yüret D. and Türe F.,** 2006. Learning Morphological Disambiguation Rules for Turkish, *Proceedings of the Human Language Technology Conference - North American Chapter of the Association for*

Computational Linguistics Annual Meeting (HLT-NAACL 2006), June 2006, New York City.

[11] **Yüret D. and Maza M.,** 2006. The Greedy Prepend Algorithm for Decision List Induction, *Proceedings of the 21st International Symposium on Computer and Information Sciences, ISCIS 2006.*

[12] **Ergenç İ.,** 2002. Konuşma Dili ve Türkçe'nin Söyleyişi, Multilingual, İstanbul.

[13] **Covington M. A.,** 1994. Natural Language Processing for Prolog Programmers, Prentice Hall, New Jersey.

[14] **Manning C.D. and Schütze H**., 1999. Foundations of Statistical Natural Language Processing, MIT Press, Massachusetts.

EK A

SÖZCÜK	TOPLAM	GÖVDE	ETİKET	SIKLIK	YÜZDE
bir	14395	bir	+Det	10542	73,23
bir	14395	bir	+Num+Card	3722	25,86
bir	14395	bir	+Adj	131	0,91
bir	14395	Bir	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
bir	14395	bir	+Adverb	0	0,00
bu	9013	bu	+Det	8374	92,91
bu	9013	bu	+Pron+Demons+A3sg+Pnon+Nom	639	7,09
de	5341	de	+Conj	4816	90,17
de	5341	de	+Verb+Pos+Imp+A2sg	525	9,83
için	4190	için	+Postp+PCNom	4190	100,00
için	4190	iç	+Noun+A3sg+P2sg+Nom	0	0,00
için	4190	iç	+Noun+A3sg+Pnon+Gen	0	0,00
için	4190	iç	+Verb+Pos+Imp+A2pl	0	0,00
ile	2769	ile	+Conj	1181	42,65
ile	2769	ile	+Postp+PCNom	1090	39,36
ile	2769	il	+Noun+A3sg+Pnon+Dat	498	17,98
çok	2559	çok	+Adj	1317	51,47
çok	2559	çok	+Adverb	980	38,30
çok	2559	çok	+Postp+PCAbl	75	2,93
çok	2559	çok	+Det	0	0,00

en	2059	en	+Adverb	2059	100,00
en	2059	en	+Noun+A3sg+Pnon+Nom	0	0,00
sonra	1855	sonra	+Postp+PCAbl	1102	59,41
sonra	1855	sonra	+Adverb	753	40,59
sonra	1855	sonra	+Noun+A3sg+Pnon+Nom	0	0,00
ama	1820	ama	+Conj	1820	100,00
ama	1820	am	+Noun+A3sg+Pnon+Dat	0	0,00
ama	1820	ama	+Adj	0	0,00
ama	1820	ama	+Noun+A3sg+Pnon+Nom	0	0,00
ne	1793	ne	+Conj	563	31,40
ne	1793	ne	+Adj	404	22,53
ne	1793	ne	+Adverb	81	4,52
ne	1793	ne	+Pron+Ques+A3sg+Pnon+Nom	0	0,00
kadar	1668	kadar	+Postp+PCDat	1051	63,01
kadar	1668	kadar	+Noun+A3sg+Pnon+Nom	617	36,99
var	1575	var	+Adj	1575	100,00
var	1575	Var	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
var	1575	var	+Verb+Pos+Imp+A2sg	0	0,00
ise	1483	is	+Noun+A3sg+Pnon+Dat	652	43,96
ise	1483	i	+Verb+Pos+Cond+A3sg	0	0,00
değil	1437	değil	+Verb+Pres+A3sg	981	68,27
değil	1437	değil	+Conj	456	31,73
değil	1437	değ	+Verb^DB+Verb+Pass+Pos+Imp+A1 g	0	0,00
yeni	1383	yeni	+Adj	1373	99,28
yeni	1383	yen	+Noun+A3sg+P3sg+Nom	6	0,43
yeni	1383	yen	+Noun+A3sg+Pnon+Acc	4	0,29
yeni	1383	Yeni	+Noun+Prop+A3sg+Pnon+Nom	0	0,00

ancak	1270	ancak	+Conj	1027	80,87
ancak	1270	ancak	+Adverb	243	19,13
diye	1219	diye	+Postp+PCNom	1219	100,00
diye	1219	de	+Verb+Pos+Opt+A3sg	0	0,00
ilk	1178	ilk	+Adj	980	83,19
ilk	1178	ilk	+Adverb	30	2,55
son	1174	son	+Adj	703	59,88
son	1174	son	+Noun+A3sg+Pnon+Nom	471	40,12
son	1174	Son	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
yıl	1061	yıl	+Noun+A3sg+Pnon+Nom	1061	100,00
yıl	1061	Yıl	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
yıl	1061	yıl	+Verb+Pos+Imp+A2sg	0	0,00
yüzde	999	yüz	+Num+Card^DB+Noun+Zero+A3sg+Pnon+Loc	981	98,20
yüzde	999	yüzde	+Noun+A3sg+Pnon+Nom	15	1,50
yüzde	999	yüz	+Noun+A3sg+Pnon+Loc	3	0,30
önce	982	önce	+Adverb	630	64,15
önce	982	önce	+Postp+PCAbl	352	35,85
önce	982	ön	+Adj^DB+Adj+AsIf	0	0,00
önce	982	ön	+Adj^DB+Adverb+Ly	0	0,00
önce	982	ön	+Noun+A3sg+Pnon+Equ	0	0,00
önce	982	önce	+Noun+A3sg+Pnon+Nom	0	0,00
göre	963	göre	+Postp+PCDat	963	100,00
göre	963	gör	+Verb+Pos+Opt+A3sg	0	0,00
o	911	o	+Det	848	93,08
o	911	o	+Pron+Demons+A3sg+Pnon+Nom	19	2,09
o	911	o	+Pron+Pers+A3sg+Pnon+Nom	0	0,00
yok	898	yok	+Adj	897	99,89

yok	898	yok	+Adverb	1	0,11
yok	898	Yok	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
karşı	859	karşı	+Noun+A3sg+Pnon+Nom	545	63,45
karşı	859	karşı	+Postp+PCDat	252	29,34
karşı	859	karşı	+Noun+A3sg+Pnon+Dat	29	3,38
karşı	859	karşı	+Adj	19	2,21
karşı	859	karşı	+Adverb	14	1,63
karşı	859	Karşı	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
ya	835	ya	+Conj	833	99,76
ya	835	ya	+Interj	2	0,24
ilgili	829	ilgili	+Noun+A3sg+Pnon+Nom	479	57,78
ilgili	829	ilgi	+Noun+A3sg+Pnon+Nom^DB+Adj+With	348	41,98
biz	753	Biz	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
biz	753	biz	+Noun+A3sg+Pnon+Nom	0	0,00
biz	753	biz	+Pron+Pers+A1pl+Pnon+Nom	0	0,00
böyle	742	böyle	+Adj	700	94,34
böyle	742	böyle	+Adverb	42	5,66
bile	717	bile	+Adverb	717	100,00
bile	717	Bile	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
bile	717	bil	+Verb+Pos+Opt+A3sg	0	0,00
bile	717	bile	+Verb+Pos+Imp+A2sg	0	0,00
ben	704	ben	+Noun+A3sg+Pnon+Nom	137	19,46
ben	704	Ben	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
ben	704	ben	+Pron+Pers+A1sg+Pnon+Nom	0	0,00
yer	703	yer	+Noun+A3sg+Pnon+Nom	695	98,86
yer	703	yer	+Verb+Pos+Imp+A2sg	6	0,85
yer	703	ye	+Verb+Pos+Aor+A3sg	2	0,28

yer	703	Yer	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
yer	703	ye	+Verb+Pos+Aor^DB+Adj+Zero	0	0,00
bin	701	bin	+Num+Card	701	100,00
bin	701	bin	+Verb+Pos+Imp+A2sg	0	0,00
bazı	677	bazı	+Adj	677	100,00
hiç	677	hiç	+Adverb	388	57,31
hiç	677	hiç	+Noun+A3sg+Pnon+Nom	289	42,69
bazı	677	baz	+Noun+A3sg+P3sg+Nom	0	0,00
bazı	677	baz	+Noun+A3sg+Pnon+Acc	0	0,00
bazı	677	bazı	+Det	0	0,00
hem	639	hem	+Conj	581	90,92
hem	639	hem	+Adverb	58	9,08
şöyle	629	şöyle	+Adverb	446	70,91
şöyle	629	şöyle	+Adj	183	29,09
mi	623	mi	+Ques+Pres+A3sg	623	100,00
mi	623	mi	+Noun+A3sg+Pnon+Nom	0	0,00
gelen	615	gel	+Verb+Pos^DB+Adj+PresPart	615	100,00
başka	615	başka	+Adj	551	89,59
başka	615	başka	+Postp+PCAbl	64	10,41
gelen	615	Gelen	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
gelen	615	gele	+Noun+A3sg+P2sg+Nom	0	0,00
şimdi	608	şimdi	+Adverb	334	54,93
şimdi	608	şimdi	+Noun+A3sg+Pnon+Nom	0	0,00
nasıl	600	nasıl	+Adverb	314	52,33
nasıl	600	nasıl	+Adj	286	47,67
sadece	590	sadece	+Adverb	590	100,00
sadece	590	sade	+Adj^DB+Adj+AsIf	0	0,00
sadece	590	sade	+Adj^DB+Adverb+Ly	0	0,00

sadece	590	sade	+Adj^DB+Noun+Zero+A3sg+Pnon+qu	0	0,00
üzerine	584	üzer	+Noun+A3sg+P3sg+Dat	551	94,35
üzerine	584	üzer	+Noun+A3sg+P2sg+Dat	33	5,65
üzerine	584	üz	+Verb+Pos+Aor^DB+Adj+Zero^DB Noun+Zero+A3sg+P2sg+Dat	0	0,00
üzerine	584	üz	+Verb+Pos+Aor^DB+Adj+Zero^DB Noun+Zero+A3sg+P3sg+Dat	0	0,00
birlikte	583	birlikte	+Postp+PCIns	435	74,61
birlikte	583	birlik	+Noun+A3sg+Pnon+Loc	118	20,24
birlikte	583	birlikte	+Adverb	30	5,15
birlikte	583	bir	+Adj^DB+Noun+Ness+A3sg+Pnon+oc	0	0,00
birlikte	583	bir	+Num+Card^DB+Noun+Ness+A3sg Pnon+Loc	0	0,00
eski	542	eski	+Adj	542	100,00
eski	542	Eski	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
eski	542	eski	+Verb+Pos+Imp+A2sg	0	0,00
artık	529	artık	+Adverb	370	69,94
artık	529	artık	+Adj	159	30,06
artık	529	Artık	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
artık	529	artık	+Noun+A3sg+Pnon+Nom	0	0,00
şu	491	şu	+Det	437	89,00
şu	491	şu	+Pron+Demons+A3sg+Pnon+Nom	54	11,00
doğru	487	doğru	+Adj	452	92,81
doğru	487	doğru	+Postp+PCDat	35	7,19
doğru	487	Doğru	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
doğru	487	doğru	+Noun+A3sg+Pnon+Nom	0	0,00
devam	483	devam	+Noun+A3sg+Pnon+Nom	482	99,79

devam	483	deva	+Noun+A3sg+P1sg+Nom	1	0,21
devam	483	Devam	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
tam	466	tam	+Adj	391	83,91
tam	466	tam	+Adverb	75	16,09
tam	466	Tam	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
neden	457	neden	+Noun+A3sg+Pnon+Nom	441	96,50
neden	457	neden	+Adverb	3	0,66
neden	457	ne	+Pron+Ques+A3sg+Pnon+Abl	0	0,00
biri	422	bir	+Num+Card^DB+Noun+Zero+A3sg+P3sg+Nom	58	13,74
biri	422	bir	+Num+Card^DB+Noun+Zero+A3sg+Pnon+Acc	5	1,18
biri	422	bir	+Adj^DB+Noun+Zero+A3sg+Pnon+Acc	1	0,24
biri	422	bir	+Adj^DB+Noun+Zero+A3sg+P3sg+om	0	0,00
biri	422	biri	+Pron+Quant+A3sg+P3sg+Nom	0	0,00
bizim	421	biz	+Noun+A3sg+P1sg+Nom	0	0,00
bizim	421	biz	+Noun+A3sg+Pnon+Nom^DB+Verb+Zero+Pres+A1sg	0	0,00
bizim	421	biz	+Pron+Pers+A1pl+Pnon+Gen	0	0,00
fazla	416	fazla	+Adj	339	81,49
fazla	416	fazla	+Adverb	54	12,98
fazla	416	fazla	+Postp+PCAbl	23	5,53
fazla	416	Fazla	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
fazla	416	faz	+Noun+A3sg+Pnon+Ins	0	0,00
üzerinde	390	üzer	+Noun+A3sg+P3sg+Loc	373	95,64
üzerinde	390	üzer	+Noun+A3sg+P2sg+Loc	17	4,36
üzerinde	390	üz	+Verb+Pos+Aor^DB+Adj+Zero^DB+Noun+Zero+A3sg+P2sg+Loc	0	0,00

üzerinde	390	üz	+Verb+Pos+Aor^DB+Adj+Zero^DB Noun+Zero+A3sg+P3sg+Loc	0	0,00
buna	360	bu	+Pron+Demons+A3sg+Pnon+Dat	360	100,00
benim	360	ben	+Noun+A3sg+P1sg+Nom	0	0,00
benim	360	ben	+Noun+A3sg+Pnon+Nom^DB+Verb Zero+Pres+A1sg	0	0,00
benim	360	ben	+Pron+Pers+A1sg+Pnon+Gen	0	0,00
benim	360	ben	+Pron+Pers+A1sg+Pnon+Nom^DB+ Verb+Zero+Pres+A1sg	0	0,00
buna	360	buna	+Verb+Pos+Imp+A2sg	0	0,00
ayrıca	356	ayrıca	+Adverb	356	100,00
ayrıca	356	ayrı	+Adj^DB+Adj+AsIf	0	0,00
ayrıca	356	ayrı	+Adj^DB+Adverb+Ly	0	0,00
ayrıca	356	ayrı	+Adj^DB+Noun+Zero+A3sg+Pnon+ qu	0	0,00
ayrıca	356	ayrıç	+Noun+A3sg+Pnon+Dat	0	0,00
altında	355	alt	+Noun+A3sg+P3sg+Loc	313	88,17
altında	355	altın	+Noun+A3sg+Pnon+Loc	42	11,83
altında	355	alt	+Noun+A3sg+P2sg+Loc	0	0,00
altında	355	altı	+Num+Card^DB+Noun+Zero+A3sg P2sg+Loc	0	0,00
eğer	341	eğer	+Conj	340	99,71
eğer	341	eğ	+Verb+Pos+Aor+A3sg	1	0,29
eğer	341	Eğer	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
eğer	341	eğ	+Verb+Pos+Aor^DB+Adj+Zero	0	0,00
eğer	341	eğer	+Noun+A3sg+Pnon+Nom	0	0,00
uzun	337	uzun	+Adj	337	100,00
uzun	337	Uzun	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
uzun	337	uz	+Noun+A3sg+P2sg+Nom	0	0,00

uzun	337	uz	+Noun+A3sg+Pnon+Gen	0	0,00
dış	323	dış	+Noun+A3sg+Pnon+Nom	165	51,08
dış	323	dış	+Adj	158	48,92
geri	312	geri	+Noun+A3sg+Pnon+Nom	203	65,06
geri	312	geri	+Noun+A3sg+Pnon+Dat	85	27,24
geri	312	geri	+Adj	19	6,09
geri	312	geri	+Adverb	5	1,60
onun	305	o	+Pron+Demons+A3sg+Pnon+Gen	283	92,79
onun	305	on	+Num+Card^DB+Noun+Zero+A3sg+Pnon+Gen	7	2,30
onun	305	Onun	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
onun	305	o	+Pron+Pers+A3sg+Pnon+Gen	0	0,00
onun	305	on	+Num+Card^DB+Noun+Zero+A3sg+P2sg+Nom	0	0,00
az	303	az	+Adj	286	94,39
az	303	az	+Postp+PCAbl	17	5,61
az	303	Az	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
az	303	az	+Adverb	0	0,00
az	303	az	+Verb+Pos+Imp+A2sg	0	0,00
açık	300	açık	+Adj	298	99,33
açık	300	açık	+Adverb	2	0,67
açık	300	Açık	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
iç	297	iç	+Noun+A3sg+Pnon+Nom	211	71,04
iç	297	iç	+Adj	86	28,96
iç	297	iç	+Verb+Pos+Imp+A2sg	0	0,00
kadın	291	kadın	+Noun+A3sg+Pnon+Nom	290	99,66
kadın	291	kadın	+Adj	1	0,34
kadın	291	kadı	+Noun+A3sg+P2sg+Nom	0	0,00
dışında	289	dış	+Noun+A3sg+P3sg+Loc	279	96,54

dışında	289	dış	+Noun+A3sg+P2sg+Loc	10	3,46
dışında	289	dışı	+Adj^DB+Noun+Zero+A3sg+P2sg+oc	0	0,00
yana	288	yan	+Noun+A3sg+Pnon+Dat	167	57,99
yana	288	yana	+Postp+PCAbl	121	42,01
yana	288	yan	+Verb+Pos+Opt+A3sg	0	0,00
önceki	282	önce	+Noun+A3sg+Pnon+Nom^DB+Adj+Rel	0	0,00
önceki	282	önce	+Postp+PCAbl^DB+Noun+Zero+A3g+Pnon+Nom^DB+Adj+Rel	0	0,00
bize	281	biz	+Noun+A3sg+Pnon+Dat	0	0,00
bize	281	biz	+Pron+Pers+A1pl+Pnon+Dat	0	0,00
bana	274	Bana	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
bana	274	ban	+Noun+A3sg+Pnon+Dat	0	0,00
bana	274	ban	+Verb+Pos+Opt+A3sg	0	0,00
bana	274	ben	+Pron+Pers+A1sg+Pnon+Dat	0	0,00
hatta	266	hatta	+Conj	143	53,76
hatta	266	hat	+Noun+A3sg+Pnon+Loc	123	46,24
hatta	266	hat	+Noun+A3sg+Pnon+Dat	0	0,00
hatta	266	hatta	+Adverb	0	0,00
dek	253	dek	+Postp+PCDat	253	100,00
dek	253	dek	+Noun+A3sg+Pnon+Nom	0	0,00
gerek	249	gerek	+Noun+A3sg+Pnon+Nom	244	97,99
gerek	249	gerek	+Conj	5	2,01
gerek	249	Gerek	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
gerek	249	gerek	+Verb+Pos+Imp+A2sg	0	0,00
orta	247	orta	+Noun+A3sg+Pnon+Nom	157	63,56
orta	247	orta	+Adj	90	36,44
orta	247	Orta	+Noun+Prop+A3sg+Pnon+Nom	0	0,00

sık	244	sık	+Adverb	192	78,69
yakın	244	yakın	+Noun+A3sg+Pnon+Nom	132	54,10
yakın	244	yakın	+Adj	112	45,90
sık	244	sık	+Adj	38	15,57
sık	244	sık	+Verb+Pos+Imp+A2sg	14	5,74
yakın	244	Yakın	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
yakın	244	yak	+Verb+Pos+Imp+A2pl	0	0,00
yakın	244	yakı	+Noun+A3sg+P2sg+Nom	0	0,00
yakın	244	yakın	+Verb+Pos+Imp+A2sg	0	0,00
kısa	243	kısa	+Adj	243	100,00
kısa	243	Kısa	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
kısa	243	kıs	+Verb+Pos+Opt+A3sg	0	0,00
halinde	236	hal	+Noun+A3sg+P3sg+Loc	221	93,64
sonunda	236	son	+Noun+A3sg+P3sg+Loc	152	64,41
sonunda	236	sonunda	+Adverb	83	35,17
halinde	236	hal	+Noun+A3sg+P2sg+Loc	15	6,36
sonunda	236	son	+Noun+A3sg+P2sg+Loc	1	0,42
onu	235	o	+Pron+Demons+A3sg+Pnon+Acc	221	94,04
yanlış	235	yanlış	+Adj	136	57,87
yanlış	235	yanlış	+Noun+A3sg+Pnon+Nom	99	42,13
onu	235	on	+Num+Card^DB+Noun+Zero+A3sg+Pnon+Acc	9	3,83
onu	235	on	+Num+Card^DB+Noun+Zero+A3sg+P3sg+Nom	4	1,70
onu	235	o	+Pron+Pers+A3sg+Pnon+Acc	0	0,00
belli	230	belli	+Adj	230	100,00
belli	230	Belli	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
belli	230	bel	+Noun+A3sg+Pnon+Nom^DB+Adj+With	0	0,00

biraz	228	biraz	+Adj	128	56,14
biraz	228	biraz	+Adverb	100	43,86
beni	228	ben	+Noun+A3sg+P3sg+Nom	92	40,35
beni	228	ben	+Noun+A3sg+Pnon+Acc	84	36,84
beni	228	ben	+Pron+Pers+A1sg+Pnon+Acc	0	0,00
ileri	217	ileri	+Noun+A3sg+Pnon+Dat	108	49,77
ileri	217	ileri	+Noun+A3sg+Pnon+Nom	103	47,47
ileri	217	ileri	+Adj	6	2,76
hala	212	hala	+Adverb	130	61,32
hala	212	hala	+Noun+A3sg+Pnon+Nom	82	38,68
hala	212	Hala	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
öyle	202	öyle	+Adj	155	76,73
öyle	202	öyle	+Adverb	47	23,27
hale	198	hal	+Noun+A3sg+Pnon+Dat	179	90,40
hale	198	hale	+Noun+A3sg+Pnon+Nom	17	8,59
hale	198	Hale	+Noun+Prop+A3sg+Pnon+Nom	2	1,01
türlü	192	türlü	+Adj	178	92,71
türlü	192	türlü	+Noun+A3sg+Pnon+Nom	8	4,17
türlü	192	tür	+Noun+A3sg+Pnon+Nom^DB+Adj+With	6	3,13
türlü	192	Türlü	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
böylece	181	böylece	+Adverb	181	100,00
onlar	181	o	+Pron+Demons+A3pl+Pnon+Nom	165	91,16
onlar	181	on	+Num+Card^DB+Noun+Zero+A3pl+Pnon+Nom	11	6,08
böylece	181	böyle	+Adj^DB+Adj+AsIf	0	0,00
böylece	181	böyle	+Adj^DB+Adverb+Ly	0	0,00
böylece	181	böyle	+Adj^DB+Noun+Zero+A3sg+Pnon+qu	0	0,00

onlar	181	o	+Pron+Pers+A3pl+Pnon+Nom	0	0,00
onlar	181	on	+Num+Card^DB+Verb+Zero+Pres+A3pl	0	0,00
üst	180	üst	+Adj	110	61,11
üst	180	üst	+Noun+A3sg+Pnon+Nom	70	38,89
bizi	175	biz	+Noun+A3sg+P3sg+Nom	0	0,00
bizi	175	biz	+Noun+A3sg+Pnon+Acc	0	0,00
bizi	175	biz	+Pron+Pers+A1pl+Pnon+Acc	0	0,00
beri	174	beri	+Postp+PCAbl	174	100,00
beri	174	beri	+Adverb	0	0,00
beri	174	beri	+Noun+A3sg+Pnon+Dat	0	0,00
beri	174	beri	+Noun+A3sg+Pnon+Nom	0	0,00
oysa	171	oysa	+Conj	161	94,15
oysa	171	oysa	+Adverb	10	5,85
oysa	171	o	+Pron+Demons+A3sg+Pnon+Nom^DB+Verb+Zero+Cond+A3sg	0	0,00
oysa	171	o	+Pron+Pers+A3sg+Pnon+Nom^DB+Verb+Zero+Cond+A3sg	0	0,00
oysa	171	oy	+Noun+A3sg+Pnon+Nom^DB+Verb+Zero+Cond+A3sg	0	0,00
oysa	171	oy	+Verb+Pos+Desr+A3sg	0	0,00
seçim	170	seçim	+Noun+A3sg+Pnon+Nom	170	100,00
önünde	170	ön	+Noun+A3sg+P3sg+Loc	163	95,88
önünde	170	ön	+Noun+A3sg+P2sg+Loc	7	4,12
seçim	170	Seçim	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
seçim	170	seçi	+Noun+A3sg+P1sg+Nom	0	0,00
boyunca	168	boyunca	+Postp+PCNom	168	100,00
önüne	168	ön	+Noun+A3sg+P3sg+Dat	163	97,02
önüne	168	ön	+Noun+A3sg+P2sg+Dat	5	2,98

boyunca	168	boy	+Noun+A3sg+P2sg+Equ	0	0,00
boyunca	168	boy	+Noun+A3sg+P3sg+Equ	0	0,00
boyunca	168	boyun	+Noun+A3sg+Pnon+Equ	0	0,00
aslında	160	aslında	+Adverb	160	100,00
aslında	160	as	+Noun+A3sg+Pnon+Nom^DB+Adj+With^DB+Noun+Zero+A3sg+P2sg+oc	0	0,00
aslında	160	asıl	+Adj^DB+Noun+Zero+A3sg+P2sg+oc	0	0,00
aslında	160	asıl	+Adj^DB+Noun+Zero+A3sg+P3sg+oc	0	0,00
amacıyla	157	amaç	+Noun+A3sg+P3sg+Ins	157	100,00
güney	157	güney	+Noun+A3sg+Pnon+Nom	134	85,35
ona	157	o	+Pron+Demons+A3sg+Pnon+Dat	133	84,71
gerçekten	157	gerçekter	+Adverb	123	78,34
gerçekten	157	gerçek	+Noun+A3sg+Pnon+Abl	33	21,02
ona	157	on	+Num+Card^DB+Noun+Zero+A3sg+Pnon+Dat	23	14,65
güney	157	güney	+Adj	12	7,64
güney	157	Güney	+Noun+Prop+A3sg+Pnon+Nom	11	7,01
amacıyla	157	ama	+Adj^DB+Noun+Agt+A3sg+Pnon+Ins	0	0,00
amacıyla	157	ama	+Noun+A3sg+Pnon+Nom^DB+Noun+Agt+A3sg+Pnon+Ins	0	0,00
ona	157	Ona	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
ona	157	o	+Pron+Pers+A3sg+Pnon+Dat	0	0,00
ona	157	ona	+Verb+Pos+Imp+A2sg	0	0,00
işte	156	iş	+Noun+A3sg+Pnon+Loc	60	38,46
işte	156	işte	+Adverb	49	31,41
başta	155	baş	+Noun+A3sg+Pnon+Loc	140	90,32

başta	155	başta	+Adverb	15	9,68
başta	155	Başta	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
bütün	152	bütün	+Adj	152	100,00
bütün	152	Bütün	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
onların	150	o	+Pron+Demons+A3pl+Pnon+Gen	133	88,67
onların	150	on	+Num+Card^DB+Noun+Zero+A3pl+Pnon+Gen	13	8,67
onların	150	o	+Pron+Pers+A3pl+Pnon+Gen	0	0,00
onların	150	on	+Num+Card^DB+Noun+Zero+A3pl+P2sg+Nom	0	0,00
yoksa	149	yok	+Adj^DB+Verb+Zero+Cond+A3sg	149	100,00
karşın	149	karşın	+Postp+PCDat	148	99,33
karşın	149	karşı	+Noun+A3sg+P2sg+Nom	1	0,67
karşın	149	karşı	+Adj^DB+Noun+Zero+A3sg+P2sg+om	0	0,00
karşın	149	karşı	+Postp+PCDat^DB+Noun+Zero+A3g+P2sg+Nom	0	0,00
yoksa	149	yoksa	+Verb+Pos+Imp+A2sg	0	0,00
yalnız	148	yalnız	+Adj	137	92,57
yalnız	148	yalnız	+Adverb	11	7,43
yalnız	148	Yalnız	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
				91155	

EK B

SÖZCÜK	TOPLAM	GÖVDE	ETİKET	SIKLIK	YÜZDE
bir	1214	Bir	+Noun+Prop+A3sg+Pnon+N m	0	0,00
bir	1214	bir	+Adj	7	0,58
bir	1214	bir	+Adverb	28	2,31
bir	1214	bir	+Det	1064	87,64
bir	1214	bir	+Num+Card	104	8,57
bu	554	bu	+Det	477	86,10
bu	554	bu	+Pron+Demons+A3sg+Pnon Nom	77	13,90
de	433	de	+Conj	432	99,77
de	433	de	+Verb+Pos+Imp+A2sg	0	0,00
o	308	o	+Det	125	40,58
o	308	o	+Pron+Demons+A3sg+Pnon Nom	28	9,09
o	308	o	+Pron+Pers+A3sg+Pnon+No	0	0,00
için	229	iç	+Noun+A3sg+P2sg+Nom	0	0,00
için	229	iç	+Noun+A3sg+Pnon+Gen	0	0,00
için	229	iç	+Verb+Pos+Imp+A2pl	0	0,00
için	229	için	+Postp+PCNom	229	100,00
ama	192	am	+Noun+A3sg+Pnon+Dat	0	0,00
ama	192	ama	+Adj	0	0,00
ama	192	ama	+Conj	192	100,00

ama	192	ama	+Noun+A3sg+Pnon+Nom	0	0,00
ben	187	Ben	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
ben	187	ben	+Noun+A3sg+Pnon+Nom	1	0,53
ben	187	ben	+Pron+Pers+A1sg+Pnon+Nom	0	0,00
diye	172	de	+Verb+Pos+Opt+A3sg	3	1,74
diye	172	diye	+Postp+PCNom	169	98,26
ne	170	ne	+Adj	10	5,88
ne	170	ne	+Adverb	18	10,59
ne	170	ne	+Conj	18	10,59
ne	170	ne	+Pron+Ques+A3sg+Pnon+Nom	0	0,00
çok	169	çok	+Adj	42	24,85
çok	169	çok	+Adverb	120	71,01
çok	169	çok	+Det	0	0,00
çok	169	çok	+Postp+PCAbl	7	4,14
kadar	148	kadar	+Noun+A3sg+Pnon+Nom	20	13,51
kadar	148	kadar	+Postp+PCDat	127	85,81
sonra	140	sonra	+Adverb	54	38,57
sonra	140	sonra	+Noun+A3sg+Pnon+Nom	0	0,00
sonra	140	sonra	+Postp+PCAbl	86	61,43
ile	116	il	+Noun+A3sg+Pnon+Dat	0	0,00
ile	116	ile	+Conj	98	84,48
ile	116	ile	+Postp+PCNom	18	15,52
en	109	en	+Adverb	108	99,08
en	109	en	+Noun+A3sg+Pnon+Nom	0	0,00
var	108	Var	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
var	108	var	+Adj	24	22,22

var	108	var	+Verb+Pos+Imp+A2sg	28	25,93
değil	101	değ	+Verb^DB+Verb+Pass+Pos+mp+A2sg	1	0,99
değil	101	değil	+Conj	36	35,64
değil	101	değil	+Verb+Pres+A3sg	64	63,37
bana	93	Bana	+Noun+Prop+A3sg+Pnon+Nm	0	0,00
bana	93	ban	+Noun+A3sg+Pnon+Dat	0	0,00
bana	93	ban	+Verb+Pos+Opt+A3sg	0	0,00
bana	93	ben	+Pron+Pers+A1sg+Pnon+Dat	0	0,00
beni	91	ben	+Noun+A3sg+P3sg+Nom	1	1,10
beni	91	ben	+Noun+A3sg+Pnon+Acc	1	1,10
beni	91	ben	+Pron+Pers+A1sg+Pnon+Acc	0	0,00
önce	81	ön	+Adj^DB+Adj+AsIf	0	0,00
önce	81	ön	+Adj^DB+Adverb+Ly	0	0,00
önce	81	ön	+Noun+A3sg+Pnon+Equ	0	0,00
önce	81	önce	+Adverb	28	34,57
önce	81	önce	+Noun+A3sg+Pnon+Nom	0	0,00
önce	81	önce	+Postp+PCAbl	53	65,43
benim	77	ben	+Noun+A3sg+P1sg+Nom	1	1,30
benim	77	ben	+Noun+A3sg+Pnon+Nom^DB+Verb+Zero+Pres+A1sg	0	0,00
benim	77	ben	+Pron+Pers+A1sg+Pnon+Gen	0	0,00
benim	77	ben	+Pron+Pers+A1sg+Pnon+Nom^DB+Verb+Zero+Pres+A1sg	0	0,00
yeni	76	Yeni	+Noun+Prop+A3sg+Pnon+Nm	0	0,00
yeni	76	yen	+Noun+A3sg+P3sg+Nom	0	0,00
yeni	76	yen	+Noun+A3sg+Pnon+Acc	0	0,00

yeni	76	yeni	+Adj	72	94,74
hiç	75	hiç	+Adverb	72	96,00
hiç	75	hiç	+Noun+A3sg+Pnon+Nom	3	4,00
onu	70	o	+Pron+Demons+A3sg+Pnon+Acc	11	15,71
onu	70	o	+Pron+Pers+A3sg+Pnon+Acc	0	0,00
onu	70	on	+Num+Card^DB+Noun+Zero+A3sg+P3sg+Nom	0	0,00
onu	70	on	+Num+Card^DB+Noun+Zero+A3sg+Pnon+Acc	0	0,00
böyle	67	böyle	+Adj	19	28,36
böyle	67	böyle	+Adverb	48	71,64
ise	66	i	+Verb+Pos+Cond+A3sg	0	0,00
ise	66	is	+Noun+A3sg+Pnon+Dat	5	7,58
mi	66	mi	+Noun+A3sg+Pnon+Nom	0	0,00
mi	66	mi	+Ques+Pres+A3sg	66	100,00
onun	65	Onun	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
onun	65	o	+Pron+Demons+A3sg+Pnon+Gen	11	16,92
onun	65	o	+Pron+Pers+A3sg+Pnon+Gen	0	0,00
onun	65	on	+Num+Card^DB+Noun+Zero+A3sg+P2sg+Nom	0	0,00
onun	65	on	+Num+Card^DB+Noun+Zero+A3sg+Pnon+Gen	0	0,00
ona	60	Ona	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
ona	60	o	+Pron+Demons+A3sg+Pnon+Dat	9	15,00
ona	60	o	+Pron+Pers+A3sg+Pnon+Dat	0	0,00
ona	60	on	+Num+Card^DB+Noun+Zero	0	0,00

			+A3sg+Pnon+Dat		
ona	60	ona	+Verb+Pos+Imp+A2sg	0	0,00
son	60	Son	+Noun+Prop+A3sg+Pnon+N m	0	0,00
son	60	son	+Adj	49	81,67
son	60	son	+Noun+A3sg+Pnon+Nom	8	13,33
yok	58	Yok	+Noun+Prop+A3sg+Pnon+N m	0	0,00
yok	58	yok	+Adj	15	25,86
yok	58	yok	+Adverb	16	27,59
başka	57	başka	+Adj	49	85,96
başka	57	başka	+Postp+PCAbl	8	14,04
nasıl	57	nasıl	+Adj	2	3,51
nasıl	57	nasıl	+Adverb	53	92,98
ya	57	ya	+Conj	47	82,46
ya	57	ya	+Interj	10	17,54
doğru	56	Doğru	+Noun+Prop+A3sg+Pnon+N m	0	0,00
doğru	56	doğru	+Adj	22	39,29
doğru	56	doğru	+Noun+A3sg+Pnon+Nom	0	0,00
doğru	56	doğru	+Postp+PCDat	31	55,36
ancak	53	ancak	+Adverb	11	20,75
ancak	53	ancak	+Conj	42	79,25
artık	53	Artık	+Noun+Prop+A3sg+Pnon+N m	0	0,00
artık	53	artık	+Adj	1	1,89
artık	53	artık	+Adverb	52	98,11
artık	53	artık	+Noun+A3sg+Pnon+Nom	0	0,00
biraz	52	biraz	+Adj	8	15,38

biraz	52	biraz	+Adverb	44	84,62
bile	51	bil	+Verb+Pos+Opt+A3sg	0	0,00
bile	51	bile	+Adverb	49	96,08
bile	51	bile	+Verb+Pos+Imp+A2sg	0	0,00
göre	51	gör	+Verb+Pos+Opt+A3sg	0	0,00
göre	51	göre	+Postp+PCDat	51	100,00
bütün	49	Bütün	+Noun+Prop+A3sg+Pnon+N m	0	0,00
bütün	49	bütün	+Adj	14	28,57
ilk	49	ilk	+Adj	37	75,51
ilk	49	ilk	+Adverb	12	24,49
kadın	46	kadı	+Noun+A3sg+P2sg+Nom	0	0,00
kadın	46	kadın	+Adj	0	0,00
kadın	46	kadın	+Noun+A3sg+Pnon+Nom	46	100,00
hem	45	hem	+Adverb	6	13,33
hem	45	hem	+Conj	39	86,67
yıl	43	Yıl	+Noun+Prop+A3sg+Pnon+N m	0	0,00
yıl	43	yıl	+Noun+A3sg+Pnon+Nom	41	95,35
yıl	43	yıl	+Verb+Pos+Imp+A2sg	2	4,65
ilgili	37	ilgi	+Noun+A3sg+Pnon+Nom^D +Adj+With	17	45,95
ilgili	37	ilgili	+Noun+A3sg+Pnon+Nom	20	54,05
birlikte	36	bir	+Adj^DB+Noun+Ness+A3sg Pnon+Loc	0	0,00
birlikte	36	bir	+Num+Card^DB+Noun+Nes +A3sg+Pnon+Loc	0	0,00
birlikte	36	birlik	+Noun+A3sg+Pnon+Loc	1	2,78
birlikte	36	birlikte	+Adverb	16	44,44

birlikte	36	birlikte	+Postp+PCIns	19	52,78
biz	36	Biz	+Noun+Prop+A3sg+Pnon+N m	0	0,00
biz	36	biz	+Noun+A3sg+Pnon+Nom	1	2,78
biz	36	biz	+Pron+Pers+A1pl+Pnon+Nor	0	0,00
uzun	35	Uzun	+Noun+Prop+A3sg+Pnon+N m	0	0,00
uzun	35	uz	+Noun+A3sg+P2sg+Nom	0	0,00
uzun	35	uz	+Noun+A3sg+Pnon+Gen	0	0,00
uzun	35	uzun	+Adj	35	100,00
biri	34	bir	+Adj^DB+Noun+Zero+A3sg P3sg+Nom	1	2,94
biri	34	bir	+Adj^DB+Noun+Zero+A3sg Pnon+Acc	1	2,94
biri	34	bir	+Num+Card^DB+Noun+Zero +A3sg+P3sg+Nom	5	14,71
biri	34	bir	+Num+Card^DB+Noun+Zero +A3sg+Pnon+Acc	4	11,76
biri	34	biri	+Pron+Quant+A3sg+P3sg+N m	0	0,00
eski	34	Eski	+Noun+Prop+A3sg+Pnon+N m	0	0,00
eski	34	eski	+Adj	33	97,06
eski	34	eski	+Verb+Pos+Imp+A2sg	1	2,94
tam	31	Tam	+Noun+Prop+A3sg+Pnon+N m	0	0,00
tam	31	tam	+Adj	14	45,16
tam	31	tam	+Adverb	17	54,84
bazı	28	baz	+Noun+A3sg+P3sg+Nom	0	0,00
bazı	28	baz	+Noun+A3sg+Pnon+Acc	0	0,00
bazı	28	bazı	+Adj	27	96,43

bazı	28	bazı	+Det	1	3,57
gelen	28	Gelen	+Noun+Prop+A3sg+Pnon+N m	0	0,00
gelen	28	gel	+Verb+Pos^DB+Adj+PresPa	28	100,00
gelen	28	gele	+Noun+A3sg+P2sg+Nom	0	0,00
şimdi	28	şimdi	+Adverb	28	100,00
şimdi	28	şimdi	+Noun+A3sg+Pnon+Nom	0	0,00
şu	27	şu	+Det	24	88,89
şu	27	şu	+Pron+Demons+A3sg+Pnon+ Nom	3	11,11
aslında	26	as	+Noun+A3sg+Pnon+Nom^D +Adj+With^DB+Noun+Zero+ A3sg+P2sg+Loc	0	0,00
aslında	26	aslında	+Adverb	26	100,00
aslında	26	asıl	+Adj^DB+Noun+Zero+A3sg+ P2sg+Loc	0	0,00
aslında	26	asıl	+Adj^DB+Noun+Zero+A3sg+ P3sg+Loc	0	0,00
öyle	26	öyle	+Adj	4	15,38
öyle	26	öyle	+Adverb	22	84,62
bin	25	bin	+Num+Card	18	72,00
bin	25	bin	+Verb+Pos+Imp+A2sg	0	0,00
bize	25	biz	+Noun+A3sg+Pnon+Dat	0	0,00
bize	25	biz	+Pron+Pers+A1pl+Pnon+Dat	0	0,00
işte	24	iş	+Noun+A3sg+Pnon+Loc	2	8,33
işte	24	işte	+Adverb	21	87,50
neden	24	ne	+Pron+Ques+A3sg+Pnon+Ab	0	0,00
neden	24	neden	+Adverb	4	16,67
neden	24	neden	+Noun+A3sg+Pnon+Nom	2	8,33
şöyle	24	şöyle	+Adj	4	16,67

şöyle	24	şöyle	+Adverb	20	83,33
yüzde	23	yüz	+Noun+A3sg+Pnon+Loc	0	0,00
yüzde	23	yüz	+Num+Card^DB+Noun+Zero+A3sg+Pnon+Loc	21	91,30
yüzde	23	yüzde	+Noun+A3sg+Pnon+Nom	2	8,70
üzerine	23	üz	+Verb+Pos+Aor^DB+Adj+Zero^DB+Noun+Zero+A3sg+P2sg+Dat	0	0,00
üzerine	23	üz	+Verb+Pos+Aor^DB+Adj+Zero^DB+Noun+Zero+A3sg+P3sg+Dat	0	0,00
üzerine	23	üzer	+Noun+A3sg+P2sg+Dat	1	4,35
üzerine	23	üzer	+Noun+A3sg+P3sg+Dat	16	69,57
az	22	Az	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
az	22	az	+Adj	8	36,36
az	22	az	+Adverb	14	63,64
az	22	az	+Postp+PCAbl	0	0,00
az	22	az	+Verb+Pos+Imp+A2sg	0	0,00
bizim	22	biz	+Noun+A3sg+P1sg+Nom	1	4,55
bizim	22	biz	+Noun+A3sg+Pnon+Nom^DB+Verb+Zero+Pres+A1sg	0	0,00
bizim	22	biz	+Pron+Pers+A1pl+Pnon+Gen	0	0,00
buna	22	bu	+Pron+Demons+A3sg+Pnon+Dat	22	100,00
buna	22	buna	+Verb+Pos+Imp+A2sg	0	0,00
sadece	22	sade	+Adj^DB+Adj+AsIf	0	0,00
sadece	22	sade	+Adj^DB+Adverb+Ly	0	0,00
sadece	22	sade	+Adj^DB+Noun+Zero+A3sg+Pnon+Equ	0	0,00
sadece	22	sadece	+Adverb	22	100,00

hatta	21	hat	+Noun+A3sg+Pnon+Dat	0	0,00
hatta	21	hat	+Noun+A3sg+Pnon+Loc	0	0,00
hatta	21	hatta	+Adverb	0	0,00
hatta	21	hatta	+Conj	21	100,00
onların	21	o	+Pron+Demons+A3pl+Pnon+Gen	3	14,29
onların	21	o	+Pron+Pers+A3pl+Pnon+Gen	0	0,00
onların	21	on	+Num+Card^DB+Noun+Zero+A3pl+P2sg+Nom	0	0,00
onların	21	on	+Num+Card^DB+Noun+Zero+A3pl+Pnon+Gen	1	4,76
karşı	20	Karşı	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
karşı	20	karşı	+Adj	1	5,00
karşı	20	karşı	+Adverb	3	15,00
karşı	20	karşı	+Noun+A3sg+Pnon+Dat	0	0,00
karşı	20	karşı	+Noun+A3sg+Pnon+Nom	1	5,00
karşı	20	karşı	+Postp+PCDat	15	75,00
üzerinde	19	üz	+Verb+Pos+Aor^DB+Adj+Zero^DB+Noun+Zero+A3sg+P2sg+Loc	0	0,00
üzerinde	19	üz	+Verb+Pos+Aor^DB+Adj+Zero^DB+Noun+Zero+A3sg+P3sg+Loc	0	0,00
üzerinde	19	üzer	+Noun+A3sg+P2sg+Loc	0	0,00
üzerinde	19	üzer	+Noun+A3sg+P3sg+Loc	17	89,47
oysa	18	o	+Pron+Demons+A3sg+Pnon+Nom^DB+Verb+Zero+Cond+A3sg	0	0,00
oysa	18	o	+Pron+Pers+A3sg+Pnon+No^DB+Verb+Zero+Cond+A3s	0	0,00

oysa	18	oy	+Noun+A3sg+Pnon+Nom^D +Verb+Zero+Cond+A3sg	0	0,00
oysa	18	oy	+Verb+Pos+Desr+A3sg	0	0,00
oysa	18	oysa	+Adverb	5	27,78
oysa	18	oysa	+Conj	13	72,22
sonunda	18	son	+Noun+A3sg+P2sg+Loc	0	0,00
sonunda	18	son	+Noun+A3sg+P3sg+Loc	2	11,11
sonunda	18	sonunda	+Adverb	16	88,89
boyunca	17	boy	+Noun+A3sg+P2sg+Equ	0	0,00
boyunca	17	boy	+Noun+A3sg+P3sg+Equ	1	5,88
boyunca	17	boyun	+Noun+A3sg+Pnon+Equ	0	0,00
boyunca	17	boyunca	+Postp+PCNom	16	94,12
geri	17	geri	+Adj	2	11,76
geri	17	geri	+Adverb	8	47,06
geri	17	geri	+Noun+A3sg+Pnon+Dat	4	23,53
geri	17	geri	+Noun+A3sg+Pnon+Nom	3	17,65
belli	16	Belli	+Noun+Prop+A3sg+Pnon+N m	0	0,00
belli	16	bel	+Noun+A3sg+Pnon+Nom^D +Adj+With	0	0,00
belli	16	belli	+Adj	15	93,75
onlar	16	o	+Pron+Demons+A3pl+Pnon+ Nom	4	25,00
onlar	16	o	+Pron+Pers+A3pl+Pnon+Nor	0	0,00
onlar	16	on	+Num+Card^DB+Noun+Zero +A3pl+Pnon+Nom	0	0,00
onlar	16	on	+Num+Card^DB+Verb+Zero Pres+A3pl	0	0,00
yanlış	16	yanlış	+Adj	9	56,25
yanlış	16	yanlış	+Noun+A3sg+Pnon+Nom	6	37,50

yoksa	15	yok	+Adj^DB+Verb+Zero+Cond+A3sg	10	66,67
yoksa	15	yoksa	+Verb+Pos+Imp+A2sg	4	26,67
bizi	14	biz	+Noun+A3sg+P3sg+Nom	0	0,00
bizi	14	biz	+Noun+A3sg+Pnon+Acc	0	0,00
bizi	14	biz	+Pron+Pers+A1pl+Pnon+Acc	0	0,00
hala	14	Hala	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
hala	14	hala	+Adverb	12	85,71
hala	14	hala	+Noun+A3sg+Pnon+Nom	2	14,29
yer	14	ye	+Verb+Pos+Aor+A3sg	2	14,29
yer	14	ye	+Verb+Pos+Aor^DB+Adj+Zero	0	0,00
yer	14	yer	+Noun+A3sg+Pnon+Nom	11	78,57
yer	14	yer	+Verb+Pos+Imp+A2sg	1	7,14
önünde	14	ön	+Noun+A3sg+P2sg+Loc	0	0,00
önünde	14	ön	+Noun+A3sg+P3sg+Loc	14	100,00
açık	13	Açık	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
açık	13	açık	+Adj	9	69,23
açık	13	açık	+Adverb	4	30,77
böylece	13	böyle	+Adj^DB+Adj+AsIf	1	7,69
böylece	13	böyle	+Adj^DB+Adverb+Ly	1	7,69
böylece	13	böyle	+Adj^DB+Noun+Zero+A3sg+Pnon+Equ	0	0,00
böylece	13	böylece	+Adverb	11	84,62
eğer	13	Eğer	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
eğer	13	eğ	+Verb+Pos+Aor+A3sg	0	0,00
eğer	13	eğ	+Verb+Pos+Aor^DB+Adj+Zero	0	0,00

			o		
eğer	13	eğer	+Conj	12	92,31
eğer	13	eğer	+Noun+A3sg+Pnon+Nom	0	0,00
fazla	13	faz	+Noun+A3sg+Pnon+Ins	0	0,00
fazla	13	fazla	+Adj	6	46,15
fazla	13	fazla	+Adverb	6	46,15
fazla	13	fazla	+Postp+PCAbl	1	7,69
gerek	13	Gerek	+Noun+Prop+A3sg+Pnon+N m	0	0,00
gerek	13	gerek	+Conj	1	7,69
gerek	13	gerek	+Noun+A3sg+Pnon+Nom	5	38,46
gerek	13	gerek	+Verb+Pos+Imp+A2sg	7	53,85
halinde	13	hal	+Noun+A3sg+P2sg+Loc	0	0,00
halinde	13	hal	+Noun+A3sg+P3sg+Loc	13	100,00
kısa	13	kıs	+Verb+Pos+Opt+A3sg	0	0,00
kısa	13	kısa	+Adj	12	92,31
türlü	13	tür	+Noun+A3sg+Pnon+Nom^D +Adj+With	0	0,00
türlü	13	türlü	+Adj	6	46,15
türlü	13	türlü	+Noun+A3sg+Pnon+Nom	3	23,08
altında	12	alt	+Noun+A3sg+P2sg+Loc	0	0,00
altında	12	alt	+Noun+A3sg+P3sg+Loc	12	100,00
altında	12	altı	+Num+Card^DB+Noun+Zero +A3sg+P2sg+Loc	0	0,00
altında	12	altın	+Noun+A3sg+Pnon+Loc	0	0,00
beri	12	beri	+Adverb	2	16,67
beri	12	beri	+Noun+A3sg+Pnon+Dat	0	0,00
beri	12	beri	+Noun+A3sg+Pnon+Nom	0	0,00
beri	12	beri	+Postp+PCAbl	10	83,33

yana	12	yan	+Noun+A3sg+Pnon+Dat	8	66,67
yana	12	yan	+Verb+Pos+Opt+A3sg	0	0,00
yana	12	yana	+Postp+PCAbl	4	33,33
önceki	12	önce	+Noun+A3sg+Pnon+Nom^D +Adj+Rel	7	58,33
önceki	12	önce	+Postp+PCAbl^DB+Noun+Z o+A3sg+Pnon+Nom^DB+Ad +Rel	0	0,00
önüne	12	ön	+Noun+A3sg+P2sg+Dat	2	16,67
önüne	12	ön	+Noun+A3sg+P3sg+Dat	10	83,33
iç	11	iç	+Adj	6	54,55
iç	11	iç	+Noun+A3sg+Pnon+Nom	4	36,36
iç	11	iç	+Verb+Pos+Imp+A2sg	1	9,09
seçim	11	Seçim	+Noun+Prop+A3sg+Pnon+N m	0	0,00
seçim	11	seçi	+Noun+A3sg+P1sg+Nom	0	0,00
seçim	11	seçim	+Noun+A3sg+Pnon+Nom	11	100,00
dışında	10	dış	+Noun+A3sg+P2sg+Loc	0	0,00
dışında	10	dış	+Noun+A3sg+P3sg+Loc	10	100,00
dışında	10	dışı	+Adj^DB+Noun+Zero+A3sg P2sg+Loc	0	0,00
yalnız	10	Yalnız	+Noun+Prop+A3sg+Pnon+N m	0	0,00
yalnız	10	yalnız	+Adj	1	10,00
yalnız	10	yalnız	+Adverb	9	90,00
gerçekten	9	gerçek	+Noun+A3sg+Pnon+Abl	0	0,00
gerçekten	9	gerçekten	+Adverb	9	100,00
yakın	9	yak	+Verb+Pos+Imp+A2pl	0	0,00
yakın	9	yakı	+Noun+A3sg+P2sg+Nom	0	0,00
yakın	9	yakın	+Adj	6	66,67

yakın	9	yakın	+Noun+A3sg+Pnon+Nom	3	33,33
yakın	9	yakın	+Verb+Pos+Imp+A2sg	0	0,00
ayrıca	8	ayrı	+Adj^DB+Adj+AsIf	0	0,00
ayrıca	8	ayrı	+Adj^DB+Adverb+Ly	0	0,00
ayrıca	8	ayrı	+Adj^DB+Noun+Zero+A3sg+Pnon+Equ	0	0,00
ayrıca	8	ayrıca	+Adverb	8	100,00
ayrıca	8	ayrıç	+Noun+A3sg+Pnon+Dat	0	0,00
orta	7	Orta	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
orta	7	orta	+Adj	6	85,71
orta	7	orta	+Noun+A3sg+Pnon+Nom	1	14,29
dek	6	dek	+Noun+A3sg+Pnon+Nom	0	0,00
dek	6	dek	+Postp+PCDat	6	100,00
dış	6	dış	+Adj	6	100,00
dış	6	dış	+Noun+A3sg+Pnon+Nom	0	0,00
hale	6	hal	+Noun+A3sg+Pnon+Dat	4	66,67
hale	6	hale	+Noun+A3sg+Pnon+Nom	2	33,33
karşın	6	karşı	+Adj^DB+Noun+Zero+A3sg+P2sg+Nom	0	0,00
karşın	6	karşı	+Noun+A3sg+P2sg+Nom	0	0,00
karşın	6	karşı	+Postp+PCDat^DB+Noun+Zero+A3sg+P2sg+Nom	0	0,00
karşın	6	karşın	+Postp+PCDat	6	100,00
üst	6	üst	+Adj	5	83,33
üst	6	üst	+Noun+A3sg+Pnon+Nom	1	16,67
sık	5	sık	+Adj	1	20,00
sık	5	sık	+Adverb	4	80,00
sık	5	sık	+Verb+Pos+Imp+A2sg	0	0,00

amacıyla	4	ama	+Adj^DB+Noun+Agt+A3sg+non+Ins	0	0,00
amacıyla	4	ama	+Noun+A3sg+Pnon+Nom^D+Noun+Agt+A3sg+Pnon+Ins	0	0,00
amacıyla	4	amaç	+Noun+A3sg+P3sg+Ins	4	100,00
başta	3	Başta	+Noun+Prop+A3sg+Pnon+Nom	0	0,00
başta	3	baş	+Noun+A3sg+Pnon+Loc	1	33,33
başta	3	başta	+Adverb	2	66,67
ileri	3	ileri	+Adj	1	33,33
ileri	3	ileri	+Noun+A3sg+Pnon+Dat	1	33,33
ileri	3	ileri	+Noun+A3sg+Pnon+Nom	1	33,33
güney	1	Güney	+Noun+Prop+A3sg+Pnon+Nom	1	100,00
güney	1	güney	+Adj	0	0,00
güney	1	güney	+Noun+A3sg+Pnon+Nom	0	0,00
				6002	

ÖZGEÇMİŞ

İlknur KARADENİZ 1999 yılında Haydarpaşa Anadolu Lisesi' nden, 2004 yılında Yeditepe Üniversitesi Bilgisayar Mühendisliği bölümünden yüksek onur ödülüyle mezun olmuştur. 2004 yılından bu yana İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği programında yüksek lisans yapmaktadır. İ.T.Ü. Doğal Dil İşleme Çalışma Grubuna üye olup bu konu üzerinde çalışmaktadır. 1 adet yurt dışı bildirisi bulunmaktadır.